THE INFLUENCE OF FOREIGN LANGUAGE BACKGROUND ON PERFORMANCE ON SELECTED INTELLIGENCE TESTS

A Thesis

Submitted to

the Faculty of Graduate Studies

University of Alberta, Edmonton.

In Partial Fulfillment

of the Requirements for the Degree

Master of Education

by

David Richard Olson

April 1962

ABSTRACT

The purpose of this study was to ascertain the effects of a foreign language background on selected intelligence tests with a view to identifying instruments which provide a valid estimate of the general intellectual ability of children with a foreign language background. The statistically derived hierarchical model of intelligence was adopted; a large g loading on a test was considered as a major criterion in the valid psychometric assessment of general intellectual potential. The other major criterion was the relative freedom of the test from any consistent bias against children with a foreign language background due to reliance on culturally acquired skills such as language. This bias was determined by comparing the mean scores on various intelligence tests for a unilingual and a bilingual group matched on socio-economic status, by correlating test scores with language background, and by an examination of the nature and content of each of the tests.

Nine promising non-verbal intelligence tests and subtests were administered to a representative Edmonton sample of 432 Grade VII children, 55 of whom were classified as bilingual on the basis of a Language Background Questionnaire. Conventional intelligence tests were compared with these non-verbal tests on the basis of factor loadings and relation-

ships with both language background and school achievement.

The three major findings for this study were:

1. Factor analytic results available for these same tests and subjects (Elley, 1961) identified three non-verbal tests with high g loadings accompanied by negligible group factor loadings.

2. Verbal tests showed significant differences between the unilingual and bilingual groups in favor of the unilingual whereas non-verbal tests showed no significant differences. Similarly, verbal tests showed a significant correlation with language background whereas non-verbal tests did not. The degree of significance is indicative of the dependence of verbal intelligence scores on language background.

3. Edmonton children with a moderate degree of foreign language background were not significantly different on either verbal or non-verbal tests from unilinguals of comperable socio-economic status. This suggests that verbal intelligence may be a positively decelerated function of foreign language background and not a linear one as suggested by most previous studies.

On the basis of eight weighted criteria set up in this study, the Raven's Standard Progressive Matrices was identified as the most useful test in the battery as a measure of the general intellectual ability of children with a foreign language background. The Cattell Test of g and the Lorge-Thorndike Figure Analogies rated almost as highly as the Raven's.

ACKNOWLEDGMENTS

The writer wishes to express sincere appreciation to Dr. R.S. MacArthur for supervising the study and for his invaluable criticisms and helpful suggestions; to Dr. J.D. Ayers and Dr. C.N.Uhl for their constructive criticisms; to Dr. W.B. Elley for the use of extensive analyzed data; and to Mrs. F. Olson for her valuable and extensive clerical work.

The writer is indebted also to the Carnegie Corporation of New York, the F.F. Reeve Foundation and the University of Alberta for their financial assistance.

TABLE OF CONTENTS

CHAPTER PAGE

CHAPTER PAGE

LIST OF TABLES

TABLE		PAGE

LIST OF FIGURES

CHAPTER I

THE GENERAL PROBLEM

> From the Neolithic Age to the Atomic Age man has culled his young in search of superior offspring... This generation, like every other, knows that the future of the world depends on encouraging the best. But who are the best? How can they be identified at an early age? How should one go about finding the best Scout in America, the best young scientist in our high schools, to best boy to go to college, the best son of a taxi-driver, the best daughter of an automobile plant employee, or the best all-around student for deferment from military service? How can we find our future leaders at an early age?
>
> (John Russell cited by McClelland et al, p.iii)

Binet's demonstration of the usefulness of an intelligence test for classification purposes began a trend which today ranks as one of the most important in applied psychology. Intelligence tests are used extensively in selection, prediction, guidance and research in such applied areas as education, industry, and the armed services.

One of the major trends in the development of intelligence testing has been the production of non-verbal tests. Until the mid 1930's most psychologists were content to measure intelligence by means of verbal problems. The limitations of such a test, however, soon became obvious. Jan Masaryck, who was to become the Secretary of Foreign Affairs for Czechoslovakia, during a childhood stay in the United States, was classified as a mental defective by a Binet test presumably

because of his lack of familiarity with the English language in which the test was administered. (Cronbach, 1960, p. 173) Conventional intelligence tests assume a certain facility with the English language, and familiarity with an urban, middle class culture, two conditions not always met in a bilingual community. If the means for assessing intelligent behavior is obstructed, the estimate of that intelligence is distorted.

The solution to the problem of a language handicap in intellectual assessment appeared to be the use of non-verbal tests.

In an early study of intelligence testing among children with a foreign language background Pintner (1923) stated:

> It seems to the writer that non-verbal tests alone are adequate for this purpose. It is inconceivable that children living in an English-speaking environment, hearing, speaking, reading nothing but English should not have a distinct advantage in tests requiring the finding of opposite words, the hunting for an appropriate analogy, the filling in of an uncompleted sentence, and the like, as compared with children who hear a foreign language at home and in many cases are required to communicate in a foreign language to some people in their environment. Such contrasting groups are very far from having had equal previous practice on the elements which go to make up the usual verbal test. (p. 292)

Originally non-verbal tests were performance tests involving the manipulation of tangible objects; more recently non-verbal tests have been constructed which resemble traditional intelligence tests in the nature of their problems, but differ

in that they use symbols other than words to convey the meaning.

Testmakers increased concern with the predictive power of intelligence tests to academic criteria constitutes a second major trend. In a more recent analysis, Cronbach (1960) classifies tests on a bipolar continuum one extreme of which he labels as "cold-blooded" predictors of school success. Empirical evidence (Elley, 1961) indicates that these superior predictors are largely academically oriented, relying heavily on vocabulary and information questions. Cronbach goes on to suggest that achievement examinations could accomplish the task of prediction even more effectively provided that no acute changes occur in the environmental situation.

At the other pole he describes tests that are designed to tap "potential" and hence are only remotely concerned with immediate and specific prediction. This type of test would attempt to minimize the influence of previous educational attainments and accentuate the underlying mental power or learning ability.

Chauncey (1958) of the Educational Testing Service has suggested that research is needed on methods of identifying the abilities of children from environments different from that of the majority. It would be in this situation that one would expect the greatest difference between "cold-blooded" predictor scores and scores of "potential".

The fundamental assumption underlying the use of this second type of instrument is what Cronbach (1960) terms "adaptive treatment" which vaguely resembles "mental orthopedics". That is, it is assumed that if an individual with high potential but a limited educational background or a cultural handicap is given appropriate adaptive treatment, his verbal intelligence and his educational achievement could be raised significantly. Support for this assumption is given by Sanchez (1934) who, by providing remedial instruction in the language arts, was able to raise the mean IQ of a group of second grade Spanish-speaking children from 72 to 100 over a two-year period. Under this adaptive program, it is plausible that the test which estimates potential will have greater predictive power than the traditional intelligence test.

If this assumption is feasible there is an immediate need for the type of intelligence test capable of estimating potential ability for children with a cultural or language handicap.

In a penetrating study of the search for talent in our society, McClelland et al (1953) point out: "To become actual, potential talent generally requires special training. That is the reason it is so important to identify potential talent early; we need to know on whom to expend the training necessary to actualize talent." (p. 197)

Similarly, MacArthur (1961) has suggested that a knowledge of the general intellectual ability of culturally-handicapped pupils in elementary and early junior high school would allow teachers to adapt instruction so as to capitalize on pupil potential.

The purpose of this study is to ascertain the effects of a foreign language background on intelligence test scores with a view to selecting economical measures of general intellectual ability which provide:

1. a valid estimate of intellectual potential, and
2. a minimal handicap to children with a foreign language background.

More specifically, this study is parallel to that of Elley in that it is an attempt to identify intelligence tests or subtests which meet the above criteria, for a sample of 55 Edmonton Grade VII children from homes in which a foreign language is spoken.

The general design involves the comparison of this bilingual group with a control group, matched for socio-economic status, on each of 12 intelligence tests. The results of the factor analysis carried out by Elley will be used to examine the nature of each of these tests.

CHAPTER II

REVIEW OF RELATED LITERATURE

I. INTRODUCTION

Bilingualism, the use of more than one language by an individual, has long been recognized as an important factor in the measurement of intelligence of foreign children. Over one hundred studies have been conducted in attempts to define the precise nature of this influence on the measurement of intelligence. It is the purpose of this paper to review critically the studies which have been conducted in this area in an effort to isolate the major factors which enter into the adequate assessment of the general intellectual ability of bilingual children. Several investigators, both early and recent, have concluded that there is an inverse linear relationship between degree of foreign language background and intelligence test score.

The problem, however, appears to be much more complex. Arsenian (1945) in a review of the literature was able to sum up the major findings in the area by saying,

> This summary points to the conclusion that bilingualism neither retards nor accelerates mental development, and that language handicap is most likely the factor responsible for the discrepancy between the performance of bilingual and monoglot children on verbal tests of intelligence. (p. 74)

More recent reviews have added little to Arsenian's

early summary; conclusions are somewhat inadequate in that they fail to provide an underlying rationale that can account for the variety of findings. The reviews of Sanchez (1932), Tireman (1941), James (1960), Manuel (1960) and especially Darcy (1953) are valuable in that they provide excellent bibliographies for the students in this area. That by Singer (1956) is outstanding in that he attempts to integrate some of the findings from various areas including psychology and linguistics.

Because of the inconclusiveness of the studies and reviews, the controversy still rages over the advantages and disadvantages of bilingualism, the optional age for second language acquisition and the virtues of "compulsory" Welsh.[1]

Since the area of bilingualism and intelligence test scores is exceedingly complex, this review will be restricted to those studies directly involving the measurement of intelligence. Although reference will be made to important theoretical considerations from other points of view, the literature in these specific branches will not be included in this review. Moreover, in the studies involving the measurements of intelligence for bilingual children, preference will be given to studies which:

---

[1] Compulsory Welsh, Times Educ. Supp., May 17, 1957, 2191:694.

1. have been made since a more recent review, that of Darcy.

2. have made an effort towards controlled experimentation,

3. throw some light on variables previously overlooked.

The corollary to this is that studies which amount to mere replication with no new findings may be relegated to a table, those weak in experimental design and statistical analysis may be excluded completely.

At the risk of mutilating the specific studies involved the findings will be selected and grouped on the basis of the important variables contributing to the intelligence test scores of bilingual children. These include: the nature of the test, cultural factors, socio-economic factors, type and quality of bilingualism. After citing the evidence, an attempt will be made to integrate the findings into workable framework with which one may interpret the intelligence test score of any bilingual group. This framework will be presented in the chapter "Theoretical Framework".

During the course of this review, "bilingual" shall be used to refer to any individual who is capable of understanding or speaking more than one language. Singer (1956) points out:

> Some people think of a bilingual as an "equilingual" a person who can perform proficiently in all aspects of both languages. But a bilingual's achievement may be limited to one aspect of a language, such as understanding, speaking, reading, writing or he may have varying degrees of ability in all these aspects. Actually, bilingualism and monolingualism can be thought of as opposite extremes of a continuum, with a continuum for each aspect of language (p. 445).

The construct "intelligence" shall refer to general intellectual ability, synonymous with abstract reasoning ability, the possession of which is fundamental for success in an academic situation. It is defined operationally as the score on an intelligence test. Other terms will be defined when they are introduced.

II. REVIEW OF THE LITERATURE

The Nature of the Test Employed

From all of the research done in this area, several of the most important early investigations, the more recent studies, and those with contrary findings were selected for a critical analysis. At the end of this section, most of the evidence will be grouped into a table. (See Table I).

Verbal versus Non-Verbal Tests. The experimental studies in this group share one basic research design, that is, they all compare the mean scores for bilingual and control groups on verbal and non-verbal intelligence tests.

One of the earliest studies to use this approach in

an attempt to find the influence of a foreign home language background on intelligence was that of Pintner and Keller in 1922. The sample included 367 children from English speaking homes and 674 children from foreign homes who were in the primary grades in Youngstown, Ohio. In an administration of a special revision of the 1916 Stanford-Binet Intelligence Scale (correlation with Stanford-Binet of .97), the mean for the English-speaking group was 92, the mean for the foreign group was 84.

From the Grade II students in one of the schools two groups consisting of 49 English-speaking and 56 foreign children were selected and compared on the Pintner Non-Language Group Test and the Binet Test. The English-speaking group scored 99 and 109 on the Binet and Pintner respectively while the foreign children scored 89 and 103 on these tests.

In a third part of this study, the authors compared the results of English and foreign children on the Stanford-Binet and an aggregate of three performance scales including the Pintner Cube Test and the Mare and Foal Test. The correlation between the Binet and Performance Scale tests for the English-speaking group was .64 whereas it was only .48 for the foreign group.

From these findings the authors conclude that children

with a foreign language background receive lower scores on the Binet-type test than on tests requiring minimum knowledge and use of English.

This study is seriously limited as far as conclusions regarding bilingualism are concerned in that differentiation between the two experimental groups was made on the basis of nationality. No quantitative measure of bilingualism was used. Perhaps the criterion of nationality was a more valid estimate of bilingualism in 1920 than it is today. Secondly, statistical treatment of the results is unsatisfactory by modern standards. Thirdly, a lack of adequate standardization of the tests, especially the Pintner, made conclusions based on its use somewhat suspect; no evidence of validity or reliability is presented.

In spite of these weaknesses, the conclusions drawn from this study have been upheld by subsequent research.

In a replication of this experiment the following year Pintner (1923) obtained comparable results. He concluded by saying,

> It may be true therefore, that for purposes of classification a verbal test is as good as a non-verbal test, because ability to get on in school requires the use of the English language. If, however, the school wishes to select the brighter foreign children for special work in English, a verbal test may no be so good. (p. 294)

Barke and Williams (1938) conducted an extensive survey in Wales in an attempt to ascertain the effects of bilingualism on intelligence. The subjects involved all the 10 and 11 year old students from five schools from the same mining district in South Wales. Two verbal tests, the Northumberland and the Scottish Intelligence tests and the Pintner Non-Language Group Test were administered. The results showed the unilingual children to be markedly superior on the verbal tests while the difference between the two groups on the Pintner Non-Language test was insignificant.

This study may be criticized on the basis of its lack of an adequate rating scale for bilingualism. Socio-economic factors are only crudely controlled. Although two of the differences are significant, it is difficult to attribute them directly to bilingualism.

Havighurst and Hilkevitch (1944) administered the Arthur Point Performance Scale to a representative sample of Southwestern American Indians aged 6 to 15 years. This test consisted of: Knox Cube, Sequin Form Board, Mare and Foal Test, Porteous Maze and Koh's Block Design. This test standardized on white children for a mean of 100 yielded I.Q.'s ranging from 115 for one group of Hopi Indians to 84 for the Ramah Indians; most of the groups clustered

around the mean for the whites.

The Kuhlmann-Anderson test as well as the Arthur Performance Scale was administered to 30 Sioux Indian children; these tests yielded means of 82.5 and 102.8 respectively. This small unrepresentative sample makes this conclusion somewhat tentative.

From this study it is observed that Indian children do about as well as white children on this performance test. Hence, a performance test of intelligence would be more valuable than a verbal intelligence test for placement and guidance of Indian children.

The chief weakness of this study is the nature of the test used. It may or may not measure the abstract intellectual ability required for academic success. The second criterion applies to the application of these findings. Since there is no assessment of language differences it is only of indirect value at this point.

In a study designed to test the effect of bilingualism upon the measurement of intelligence of preschool children, Darcy (1946) selected 212 children of Italian parentage and divided them into a bilingual and monolingual group on the basis of a questionnaire. The groups were matched on age, occupational level of father and sex.

The 1937 revision of the Stanford-Binet and the Atkins

Object-Fitting Test, Form A, both individual tests, were administered to both groups.

The results permitted the author to conclude that there is a significant difference between the groups on the Stanford-Binet, in favor of the monolinguals. On the Performance test, a significant difference was found between the groups in favor of the bilinguals. Moreover, the correlation between the two tests was so low that she concluded the tests were measuring different functions.

The findings also lead to the conclusion that the bilingual subjects of this investigation suffered from a language handicap in their performance on the Stanford-Binet.

This study has an outstanding experimental design. It is limited only to the extent of the unreliability of the tests at this age level, and in that the Atkin's Performance test appears to be unrelated to abstract intellectual ability.

Carlson and Henderson (1950) found that Mexican children rate significantly lower on both verbal and non-verbal tests. Although they present the problem and theoretical basis quite adequately, the research design fails completely to control the intervening variables. The sample selection, failure to quantify or rate bilingualism, method

of matching groups for socio-economic status and the use of a conglomeration of intelligence tests, make the results suspect.

Darcy (1952) studied the performance of 235 bilingual Puerto Rican children on the Pintner General Ability, Verbal Series and the Pintner General Ability, Non-Language Series. Bilingual status was assessed on two occasions to ensure its reliability.

The results showed the bilingual children significantly superior on the Non-Language Test than on the Language Tests. She further suggests that since the two tests correlate .58 they are measuring the same thing to a marked extent.

Although the first part of the study is sound, the intrepretation of the correlation is somewhat tenuous. A more adequate technique would be that of factor analysis.

Jones (1953), in a study designed primarily to find the influence of reading ability in English on scores of Welsh-speaking, bilingual children on an English intelligence test, administered the Jenkins' Non-Verbal and the Moray House Intelligence Test, to two random groups of children age 10 to 12 from an English speaking and a Welsh speaking area of Caernarvonshire. A questionnaire resembling that of Hoffman (1935) served to substantiate the division of these

children into monoglot and bilingual groups.

The results showed no significant differences between the two groups on the Jenkins Non-Language Test in either means or variability. On the other hand, he found a significant difference of 10.63 points on the verbal intelligence test, in favor of the monolinguals.

Jones then attempted to find out if the differences in verbal I.Q. remained significant if the two groups (already equated on non-verbal intelligence) were also equated in reading ability as measured by the Schonell Silent Reading Test. By means of analysis of covariance he found that the superiority of the monoglot group was reduced from 10.63 to 6.27 points but remained highly significant.

The superiority of the monoglot group on the verbal intelligence test even when matched on non-language intelligence and reading ability led Jones to hypothesize that perhaps this difference is due to the inability of the bilinguals to "think in English". To further examine this hypothesis, Jones administered a Welsh translation of a verbal test and found the mean similar to that on the non-verbal test.

Jones findings are virtually unassailable. His assertion about the inability of the Welsh children to

think in English remains highly tentative.

More recently Jones (1959) has summarized a series of studies similar to the one discussed above. Two major conclusions emerge. First, I.Q.'s obtained from the Verbal group test in English cannot be regarded as valid assessments in the case of these Welsh speaking children, and secondly, bilingualism as such need not have an adverse effect on performance in a non-verbal test of intelligence.

Altus (1953) suggested that the mean WISC Performance I.Q. would serve as a good predictor of verbal I.Q. except where some retarding influence such as bilingualism was in operation. She then tested the patterning on the Wechsler Intelligence Scale for Children for two groups of dull school children, one group of bilinguals of Mexican descent and a control group equated on age, sex and Performance I.Q. She found a high, significant difference of 17 points in favor of the unilingual group on the verbal I.Q. Test. On the individual subtests, the unilingual group scored significantly better than the bilingual group on all the verbal subtests, and not significantly different on the performance subtests. She concludes:

Although the bilingual group is thus within

> the psychometrically retarded range in verbal (English) skills, much of this retardation is apparently a linguistic one and does not reflect what the child would probably have done with training from birth in only one language - as evidenced more nearly, though of course imperfectly, by the performance score. (p.244)

When viewing the findings of Altus, one must remember that her sample was not representative of the normal population; it is highly positively skewed. Although such variables of socio-economic status and cultured influences were not controlled, the relationships found appear reliable.

Reid (1955) and Coull (1956) in an Alberta study of 131 bilingual school children and 200 randomly chosen English-speaking children in Grade VII were compared on the California Test of Mental Maturity (Short Form) and the California Language Achievement Test. The unilinguals scored significantly superior to the bilinguals on both tests. From these results the authors conclude that the linguistic advantage enjoyed by English monoglot pupils may account for their superiority over the unilinguals on the C.T.M.M.

The propriety of the method of selection of the control group is highly suspect. Important variables as socio-economic status may account for these findings. However, these findings concur with those of Jones regarding the in-

fluence of language ability on intelligence tests scores.

Ellis (1957) in a study designed to assess the effectiveness of culture-free tests in measuring the intellectual characteristics of German immigrants to Canada compared 50 immigrant children between 10 and 15 years of age with 50 second-generation Canadians on the Raven's Progressive Matrices, the Performance Scale of the Wechsler-Bellevue Intelligence Test and the I PAT Test of 'g'. The Begabungstest B-1 was administered to the immigrant group. He concluded that the I PAT was the most satisfactory of the group tests but that it was inferior to individual performance tests and to the test in the German language. The available abstract of this study fails, however, to point out the criterion for "satisfactory" although it is apparently the highest mean score. Tests with the highest scores may not be measuring general intellectual ability.

He also suggests: "This study reveals the possibility of using for inter-racial comparisons a scale based on American culture, providing all verbal elements are eliminated".

Kittell (1959) in a recent study in California compared bilingual and unilingual children on the California Test of Mental Maturity (Short Form). The 42 subjects in-

cluded all children in the high third grade over a four year period. The control group consisted of 41 children in one class in one semester, that is, one class out of eight, chosen at random. The language rating was accomplished by examining the registration form completed by the parents; 19 percent of the experimental group was Chinese.

The author found no significant differences between the verbal or non-verbal section of the CTMM for the bilingual group. The unilingual group scored significantly higher on the verbal sub-test.

The samples were then divided into sub-groups on the basis of occupational ratings of the parents and compared on the CTMM. No significant differences were observed. Hence, it was concluded that socio-economic status did not effect CTMM scores in this study.

The high scores on the CTMM by the bilinguals indicates that this group is not typical of the normal population. The small sample, the method by which it was selected as well as the selection procedure for the control group may have affected these results. Since the sample is small, high socio-economic status of the whole group may account for the unexpected high scores. Insufficient evidence is given to reconcile this study with its more

adequate predecessors. In line with other studies, however, is the fact that the unilinguals scored significantly higher than the bilinguals on the verbal but not higher on the non-language subtest.

Lewis (1959) carried out a study using only the Jenkins Non-Verbal Scale of Mental Ability to compare groups of varying degrees of bilingual ability. Using a language questionnaire similar to that of Hoffman, (1935) he selected two groups for comparison, those who spoke Welsh in their homes always and those who never did; both groups spoke English in schools, therefore, the first was bilingual, the second unilingual. The two intermediate groups represented varying degrees of bilingualism. Teachers' ratings coincided with the questionnaire results to a great extent,

Each group consisted of from 88 to 99 10-year old children from sixteen primary schools in Wales. Instructions in Welsh were given to those students who preferred them.

Lewis found that the groups were not equal in Non-Verbal test performance, "the order of decreasing 'Welshness' of background, being also one of increasing mean score", the overall differences being significant at the 5 percent level.

Lewis concluded that these results "reflect a real difference between monoglot and bilingual children in performance on the test". He then attempted to account for these

differences by hypothesizing that the bilingual child may be slower in thinking because of the interplay of the language and thus be penalized on timed tests.

Lewis points out that his conclusions may be inaccurate in that although parental occupation did not differ appreciably between the groups, a more elaborate examination may have been shown some differences. Secondly, the "Increasing Welshness" groups are associated with rural areas to a certain extent, this could have influenced the test results. Finally, although the Jenkins is reported to be Non-Language, it has been shown to have a small loading on 'v'.

Jones has severely criticized Lewis's study mainly on the basis of the use made by Lewis of the results of his language questionnaire. He feels that the instrument is too unreliable for such a precise dichotomy. He also emphasizes the rural-urban differences in the group.(Jones,1960)

It is the impression of this writer that Lewis is his own best critic. The rural-urban differences plus the nature of the questionnaire make the conclusions provocative if tentative.

A recent study in this area is that of Levinson (1960) who studied the verbal and performance ability of monolingual and bilingual native born, Jewish, preschool children of

traditional parentage.

The experiment involved the administering of the Revised Stanford-Binet, the Wechler Intelligence Scale for Children and the Colored Progressive Matrices to 65 girls and 63 boys upon application for admission to a Hebrew speaking school. The bilingual and unilingual groups, so classified by answers on registration forms, were matched on number per sex and roughly on socio-economic status. Cultural influences were nominal due to the type of individuals attracted to such a school.

Upon comparison of means for the two groups, the monolingual boys were superior to the bilingual boys on 11 out of 18 tests and sub-tests. None of these differences were significant. The monolingual girls were superior on 15 out of 18 tests and significantly superior on the WISC, both Performance and Verbal subtests and the Stanford-Binet. No differences emerged on the Raven's.

The study is well designed and of the type required to throw light on the influence of bilingualism with other variables held constant. It is limited in that the bilingualism was not quantified; we have no idea of the degree of English used in the home. Secondly, the study is of limited generalizations in that it was performed at the preschool level as well as the fact that intelligence

tests are quite unreliable at preschool ages.

Translations of Verbal Tests In this section the studies which have used a translation of an intelligence test into the second language and compared the results with those from a test in English will be considered.

Mitchell (1937) carried out a study to determine if Spanish speaking children do as well on an intelligence test administered in the English language as they would on a similar test administered in their native language. The Otis Group Intelligence Test was administered to 236 Spanish-speaking children in Grades I through III in both Spanish and English. Although this is a Non-Verbal Test, the directions given in Spanish produced an increase of 9.28 points over the English administration for the three grades combined. Mitchell concluded that bilingual children suffered a general language handicap if tested in English.

In interpreting this finding it would be useful to know the extent of the English influence on the children. A bilingualism schedule would be useful on this context. Secondly, the Spanish translation may have been more suggestive of answers than the English version. However, the influence on a non-verbal test should be minimal.

In a more recent study, Anastasi and Cordova (1953) administered two parallel forms of the Cattell "Culture-

Free" test, one in Spanish the other in English, to 108 Puerto Rican children approximately 13 years of age. The tests correlated .84. The subjects were all rated as bilingual by a Hoffman-type language rating scale. The group was divided into two with the first group taking first the English then the Spanish form while the second group took first the Spanish then the English form.

Both groups did better on the second test no matter if it was in Spanish or English. From this the authors conclude that the language in which the test was administered had no significant effect on performance, the children's mastery of either language being inadequate.

The design of this study is a good one. It could be replicated using a little more care in controlling the socio-economic and bilingual factors and using validated and standardized instruments of both a verbal and non-verbal nature.

Finally, Keston and Jiminez (1954), in a study in Albuquerque, administered the two versions of the 1937 Revision of the Stanford-Binet, one in English and one in Spanish to 50 Spanish-American Grade IV children. The Spanish version was imported from Spain and correlated with the English version .37.

It was found that their sample performed significantly

better on the English version than the Spanish version of the test. The authors conclude that both languages present difficulty to the Mexican-decent children, however, they do better on the language of their formal education.

Sanchez (1932) nearly 30 years ago reviewed the literature on the performance of Spanish speaking Americans on intelligence tests. He suggested that: "Students of linguistic difficulties of these children have been handicapped by difficulties in measurement. Translated tests are at best doubtful value." (p. 556)

In these Spanish translation experiments, the findings can be reconciled by consideration of an unspecified variable, that of the degree and type of bilingualism. In the first study, that of Mitchell, the children were more Spanish then English. However, 20 years later, Keston and Jiminez found them more adequate in English then in Spanish. This factor will be discussed more fully in a later section of this paper.

Barke and Williams (1938) in a second part of the study previously reviewed, administered a Welsh translation of the Northumberland verbal test along with the verbal tests in English and the Pintner Non-Language intelligence test. The bilinguals were found to be distinctly inferior when measured by a verbal test in their second language and also

distinctly inferior when tested in their mother tongue. The authors concluded that in neither language do they have the vocabulary equal to the monoglot.

This study was replicated by Stark (1940) who administered two verbal intelligence tests and one non-verbal test to 271 bilingual children and 271 monoglots aged 10 to 12 in nine schools in Dublin.

An Irish translation was made of the Dawson Mental Test, Form B which is reported to correlate .84 with the English version. The results showed a significant difference of five points in favor of the monoglot group at ages 11 and 12 with no differences at age 10.

Next, 104 monolinguals and 65 bilinguals chosen at random from the previous groups were compared on the Dawson Mental Test, Form A, administered in English to both groups. At ages 10 and 11 the bilingual group was significantly higher than the monoglot group.

Finally, the Passalong Test, an individual performance test, was administered to 41 monoglots and 41 bilinguals. No significant differences were found.

Stark concluded that early acquisition of a second language may reinforce the first. Stated more generally, children with a high innate predisposition for verbal facility may find a second language beneficial to their mental

development.

It should be noted that neither of these conclusions are supported by any other studies. In order to rationally account for his data one would need to have a knowledge of the degree of bilingualism and the socio-economic status. Nevertheless, it appears that for his sample the nonverbal scores in the English administration were higher than when administered in the home language.

Jones (1953) in a study reported above administered a Welsh translation of Jenkins' Cotswold Mental Ability Test to the group who had previously taken a verbal and a non-verbal test in English. The results showed the bilinguals do as well on this verbal Welsh translation as on the non-verbal test, that is, significantly better than they had done on the English verbal test.

From this Jones concluded bilinguals find it difficult to "think in English". It is impossible to assess the accuracy of this conclusion without more evidence of the nature and adequacy of the translation; Jones provides no details. Secondly, the degree of Welshness of this group, although unspecified, may have been very high; this could account for this finding. Type of bilingualism would be worth considering.

It should be noted that the translation of an in-

telligence test from one language to another, as has been done in the above studies is an extremely difficult task. In verbal tests especially, shades of meaning will undoubtedly change, clues may be given or denied, culturally familiar questions in one environment may be foreign to another.

In an attempt to show that their translation is accurate some of the writers show concurrent validity, that is, they show a correlation with the English equivalent. However, this approach is automatically limited; if the correlation approaches 1.00 the translated test is discriminating between bilinguals and unilinguals as markedly as the original test it was designed to replace. If, on the other hand, the correlation drops, the test is assumed to be invalid and not suitable for its task. The only solution would be the construction of an intelligence test, standardized on the foreign speaking population, which showed construct validity. Then, subsequent translations could show concurrent validity with this standardized measure.

The difference between the scores on an English version and a mother-language version (assuming that the tests are of similar nature) is likely due to the degree or type of bilingualism. This aspect will be discussed later.

<u>Conclusion</u>. From this review of the literature it is obvious that one important variable in the measurement of intelligence of bilingual children is the nature of the test.

(i) The verbal ability of bilinguals tends to be more limited in both languages than that of unilinguals of equal age, environment and intellectual potential. The corollory of this is that tests requiring verbal facility tend to differentiate between unilinguals and bilinguals.

(ii) Tests of a non-verbal nature which measure abstract reasoning ability or have a high factorial loading on g tend not to discriminate between unilingual and bilingual children of comparable age and environment regardless of the language of administration.

(iii) Tests requiring performance skills or specific mental abilities may handicap or bonus the bilingual child. The validity of these instruments as measures of general intelligence must be seriously questioned. There is some doubt, for instance, that the Pintner is measuring the general abstract intelligence required for scholastic achievement. Secondly, the earlier writers fail to report reliability. If this value is low, one cannot expect consistent results to emerge from replications of experiments.

(iv) Translated tests fit the pattern for the parent

TABLE I

SUMMARY OF STUDIES IN WHICH A BILINGUAL AND A CONTROL WERE COMPARED ON VARIOUS INTELLIGENCE TESTS

Student	Date	Place	Sample	Measures	Results: Exp.Group Biling.	Results: Control Uniling.	Conclusion
Pintner and Keller	1922	Urban, Ohio	49 Eng. 56 Foreign Grade 2	Binet	89	99	Children with foreign language background get lower scores on Binet-type test than on Non-language tests.
				Pintner (N.-L.)	103	109	
				Binet Aggregate Pref.	r = .48	r = .64	
Pintner	1923	N.Y.C.	Gr.3 & 4	National Intell.	8 - 7	9 - 10	Use caution in drawing conclusions from verbal intell. tests for bilinguals.
				Pintner Non-L.	9 - 4	9 - 4	
Colvin and Allen	1923	Providence, R.I.	Gr. 5-8 Italian	Stanford-Binet	91	92	Lack of verbal facility may significantly lower verbal I.Q. scores.
				National Intell.	76	85	
Jamieson and Sandiford	1928	Ontario	717 Indian Children	National Intell.	76	82	Indian children fall below white norms, especially on verbal tests.
				Pintner Non-L.	93	100	
				Pintner-Peterson	96	91	
Pintner and Arsenian	1937	N.Y.C.	Gr. 6 & 7 Jewish	Pintner Lang.	102	104	No significant difference between groups or tests.
				Pintner Non-L.	102	103	
Barke and Williams	1938	Wales	Large 10½-11½ years	Scottish Verb.	85	101	Bilinguals markedly inferior on verbal tests but equal on non-verbal.
				Northumberland	82	93	
				Pintner Non-L.	88	88	
Havighurst and Hilkevitch	1944	Western U.S.	670 Indians 6-15 years	Arthur Point Performance Scale	100		Indians do as well as whites on this test.

TABLE I (continued)

Student	Date	Place	Sample	Measures	Results: Exp. Group Biling.	Results: Control Uniling.	Conclusion
Darcy	1946	N.Y.C.	212 3-4-5 yr. olds.	Stanford-Binet	91	99	Bilingual suffer verbal handicap. Perf. test measure Diff. function
				Atkins Perf.	98	89	
Carlson and Henderson	1950	Southwest U.S.	115 Mexicans 105 Control	CTMM Lang.			Mexican children rate significantly lower on both Verb. and Non-Lang. tests.
				CTMM Non-Lang.			
Darcy	1952	N.Y.C.	235 Puerto Ricans Gr.5 and 6	Pintner Verbal	80		Puerto Rician children rate significantly lower on verbal intell. tests.
				Pintner Non-Lang.	88		
Jones	1953	Wales	64 Biling. 51 Uni.	Verbal	91	102	No. sign. diff. in non-verbal. Sign. diff. in verbal test.
				Jenkins Non-Lang.			
Atlas	1953	U.S.	Dull Mexican, small sample	WISC Verbal	72	89	All verbal subtests show significant diff. at .01 level
				Performance	84	86	
Reid	1955	Alberta	131 Biling. 200 Eng.	CTMM Total	95	105	Linguistic advantage enjoyed by Eng. monoglot.
				Cal. Lang. Test	63	69	
Cooper	1958	Guam	Gr.V, 164 children	CTMM Lang.	81		Guam children score higher on non-lang. tests.
				Non-Verbal	88		
Kittell	1959	Calif.	42 Biling.	CTMM Lang.	108	114	Bilinguals do as well on either verbal or non-verbal tests.
				Non-Lang.	109	108	
Lewis	1959	Wales	375-10 yr.	Jenkins	33	41	Increasing degree of Welshness involves decreasing score on non-verbal intell. test.
Levinson	1960	U.S.	65 Jewish preschool girls	WISC Verbal	96	106	Jewish preschool girls rated sign. higher on all tests except the Raven's.
				Performance	99	109	
				Stanford-Binet	110	120	
				Ravens	15	15	

verbal or non-verbal test. Performance on verbal tests in any language will depend on the type and degree of bilingualism.

Cultural Factors

For the purpose of this paper culture is defined as the complex of beliefs, attitudes, traditions, behavior patterns and traits, which are the stereotyped possession of an ethnic ingroup.

In this section the results of intelligence tests will be studied in reference to three ethnic subcultures. Following this, studies giving evidence of a theoretical formulation regarding cultural influences will be considered.

In an excellent article, Soffiette (1955), attempts to differentiate between bilingualism and biculturalism. He suggests that many of the previous studies directed towards bilingualism have confused or masked the findings by leaving uncontrolled the influence of biculturalism. He concludes that the retardation ascribed to bilingualism is due rather to biculturalism. "It is a conflict between ways of life, beliefs, customs, value systems and not necessarily one between language systems." (p.225)

If this is the case, the study of the effect of bilingualism could better be studied employing research designs

similar to that of Levinson (1960) and that of Jamieson and Sandiford (1928) in which a bilingual group and a unilingual group arising from the same cultural background were compared.

The influence which various cultural patterns have on intelligence test scores can only be inferred from a cross-sectional review of several studies involved with different cultural groups.

A cultural influence can be inferred from the findings of Jamieson and Sandiford (1928) in which two groups of Southern Ontario Indian children, one bilingual and the other unilingual, were compared on three intelligence tests. On the National Intelligence Test, a highly verbal instrument, both experimental and control groups fell below the mean. Since the unilingual group fell below the mean, some variable other than bilingualism appears to be operating; this factor is likely a cultural one. Since the bilingual group fell below the unilingual group this latter difference seems to be attributable to bilingualism. Because of the lack of control over such variables as socio-economic rating and rural-urban differences and the inadequacy of the statistical analysis these conclusions must remain highly tenative. The non-verbal Pintner tests failed to show a consistent pattern.

The Havighurst and Hilkevitch (1944) study reported above showed that although Indian children may score significantly lower on a verbal test of intelligence, they tend to score somewhat near the American mean on a performance scale.

From these Indian studies, it appears that a cultural factor enters particularly in verbal tests and in some groups on non-verbal tests.

A cultural influence operating in a positive direction may be inferred from the study of Levinson (1960), in which Jewish children scored above the norms for the general population. Bilingual children, however, tended to score below the unilingual group. The general superiority may be due to cultural influence or to socio-economic status.

Murdock and Maddow (1928) compared three groups of Grade VII Jewish girls dichotomized on the basis of the amount of English spoken in the home on the Otis (verbal) and International (non-language) intelligence tests. Although he found increasing intelligence scores were related to a larger amount of English spoken in the home, the tests showed no difference in their ability to differentiate between the groups.

Similarly, Pintner and Arsenian (1937) studied the effects of varying degrees of bilingualism of ten to twelve year old Jewish children on the Pintner Verbal and Pintner

Non-Language intelligence tests. To provide further control, the groups of high and low bilinguals were matched on socio-economic status.

The authors found no significant difference between the correlations of bilingualism with the Pintner verbal and with the Pintner non-language test scores, for the high and low bilingual groups. They conclude,

> Our results indicate therefore that Jewish bilingual children of the sixth and seventh grades born in this country have acquired the English language sufficiently well not to be handicapped in their performance in a group intelligence test of the usual type. (p. 259)

Spoerl (1943) found no I.Q. differences between mono-lingual and bilingual Jewish college students. Perhaps selection factors influenced this result.

Intelligence test profiles for Jewish children as have been indicated in the four above reviews, are atypical. The divergence of this group may be attributed to some cultural factor, such as high achievement need as a compensation for lack of social recognition (Clark, 1949); the general verbal superiority and the normal non-verbal ratings fit this hypothesis. The cultural and personality dynamics, however, are beyond the scope of this paper.

The other factor which contributes to an understanding of these findings is age and education. Although Levinson

(1960) found the Jewish bilingual preschoolers to be widely different on verbal and performance tests, this difference seems to have disappeared by Grade VII according to the other two studies.

A great deal of work has been done on the Spanish-speaking children of the southern United States; Sanchez (1932) concluded from his review of the literature that:

> .....it is apparent that in the education of Spanish speaking children we have a combination of practically all the factors which among other groups have been shown to impair the value of test results. (p. 556)

This view is largely substantiated by the more recent research. Mitchell (1937) showed that his Spanish-speaking bilinguals did poorly on English verbal tests. Altus (1953) showed that her group did more poorly on the verbal part of the WISC than on the performance part, however, they were significantly below the mean on both subtests. Carlson and Henderson's (1950) findings as well as those of Keston and Jiminez (1954) support those of Altus.

It appears that something more than sheer bilingualism is operating in the reduction of test scores. It is hypothesized that one influence is that of cultural background.

It also remains an untested hypothesis that some of the differences between English and English-Welsh bilinguals is due to cultural influences. This may be envoked through

cultural content of test items or through implicit reaction to testing situations.

Jones (1953) in a previously reported study, found a significant difference of 10.63 points between bilinguals and unilinguals on verbal intelligence. By analysis of variance he was able to parcel out the effect of reading ability as a factor in these test scores. However, even when this was done, there remained a significant difference between the two groups. Whether this final difference is due to some phase of bilingualism thus far overlooked or to cultural factors, is not known.

Included in this cultural factor would be attitudes and biases toward the language and customs of the national society. It is highly probable that prejudice against the English language causes a decline in the willingness to learn English, to use English and perhaps even to co-operate in an education system. Johnson (1951) has shown that people who either know a great deal or very little about the Anglo culture showed less bias toward it than those with a medium knowledge of this culture. That this bias has emotional concomitants is shown by Spoerl (1943) who found that bilingual students tend to be more emotionally maladjusted than the control. He attributes this to cultural conflicts between native born children and their foreign born parents.

Gardner and Lambert (1959) have shown the importance of the integrative motive, that is, interest in meeting with and understanding more about members of the linguistic outgroup, in the acquisition of the second language. Lambert, et al (1960) have substantiated this by showing that those with higher bilingual scores had a more favorable attitude to the second language.

It is possible to account to some extent for the differences between the cultural groups by this variable.

It is also possible that each various sub-culture as well as sharing certain general influences such as personal maladjustment to a new culture, has its own syndrome of cultural influences the nature and pattern of which can only be ascertained by continued research.

In interpreting the results of the intelligence test scores of bilinguals, one must take cognizance of the fact that cultural sets may influence the degree of bilingualism, the emotional reaction of the child to the school and testing situation, and the nature of responses on the test itself. For an adequate study of bilingualism, the experimental groups should be equated on this variable.

Socio-Economic Status

It has been suggested above that several apparently incompatable findings can be reconciled on the basis of un-

controlled socio-economic status.

Singer (1956) points out that importance of rural-urban differences. The scores on English tests tend to decrease from urban to rural areas.

Jones (1960) has made a critical review of three of the Welsh-English bilingual studies in the light of socio-economic variables. He criticized severely the conclusion that bilingualism causes a decrease in non-verbal test scores. The differences, he maintained, may be attributed to rural-urban differences and to parental-occupation level.

Socio-economic factors may have no simple relationship with intellectual factors as is pointed out by McCarthy (1954). She suggests that most studies of bilingualism are complicated by socio-economic status, for most bilingual children come from either highly cultural homes of the upper social level where the language is being preserved for cultural reasons, or from lower socio-economic levels where the parents have not been sufficiently intellectual to acquire the new language. Immigrants often remain in a lower socio-economic bracket than they would be found in their native countries because their language handicap has necessitated their remaining at manual occupations rather than undertaking more verbal and intellectual tasks.

Although cultural factors have been subsumed under socio-economic factors in most studies, it has also been shown repeatedly that decreasing socio-economic level is related to decreasing intelligence test scores, particularly on tests of a verbal nature.

Type and Degree of Bilingualism

Degree of bilingualism was found to be an important variable in the measurement of intelligence as early as 1928. (Murdock et al) Hoffman (1935) developed a questionnaire in 1935 which has served as the basis for classifying degree of bilingualism until the present. However, after making a careful review of the literature, Sr. Mary Andrew Hartmann (1961, p. 74) concluded, "The assumption that the degree of bilingualism corresponds to the degree of bilingual background can hardly be justified". The differences in bilingualism perhaps involves more than a merely quantitative one.

It appears that in regard to rating bilingualism there are two variables involved, the nature of the bilingualism and the degree or extent of bilingualism. First, let us consider the type or nature of bilingualism.

Osgood (1954) laid the foundation for theoretical and empirical study with his conception of compound and co-

ordinate bilingualism.

A compound bilingual system, in which one would have a tendency to confuse the two languages, would arise when one language is learned from a previous knowledge of another language, for example, in a school situation. It may also arise when the parents of the child speak two languages more or less interchangeably in the same situation. Under this system one would anticipate interference of the two language systems.

The coordinate bilingual, the "true" bilingual, is found when the child has learned one language in one setting, in which it was consistently used, while the other was consistently used in another situation. For instance, a child whose parents speak a pure foreign language consistently in the home, while peers and school teachers speak English at play and at school, would be a coordinate bilingual.

Lambert and Fillenbaum (1959) have provided experimental evidence for this theoretical formulation. Using several polyglot aphasics, they were able to find a direct relationship between the functional dependence or independence of the polyglot's languages and the situation in which they had been learned.

Pertaining to intelligence test functioning, the compound bilingual is expected to show more confusion and interference between the languages than the coordinate. As a

result , he will be expected to rate significantly lower than the coordinate bilingual on a verbal test of intelligence. Perhaps we could even hypothesize that thinking processes will show more interference in which case we would also expect the compound to rate lower on a non-verbal test.

If bilingualism is distributed bimodally, a new type of language rating is needed to supplement that of Hoffman.

Secondly, the degrees of bilingualism appears to be important according to the findings of Jones (1953), Lewis (1959) and others. Degree of bilingualism refers to the relative strength of the two languages involved. The higher the degree of foreign language influence, the better the verbal intelligence test scores in that language; lower foreign language scores would be related to superior scores on the English translation. With increasing age and education, the verbal handicap tends to disappear as has been shown in the Jewish studies above. At a preschool age, bilinguals were reported to be markedly inferior in verbal subtest, at the end of elementary school no differences were observed. One must take as extremely tentative this conclusion that continued education decreases the language handicap as these studies were not longitudial. These are quoted because it appears reasonable to assert that continuing education will make one more proficient in the language

of formal instruction. Terman's (1918) study of vocabulary lists with bilinguals gives indirect support for this conclusion. He found that for the bilingual student, vocabulary age is lower than mental age at Grade II and IV whereas no difference exists at Grades VI or VII. (His MA was based on a highly verbal test, however.)

To summarize, foreign language background must not only be assessed in terms of quantity of foreign language background but also in terms of the type or quality of bilingualism. Coordinate bilingualism should have a minimal influence of mental development. Degree of bilingualism probably does not effect mental development, but will have a marked effect on the assessment of intelligence by conventional tests.

The conclusions of this entire review of the literature will be presented more formally in the chapter "Theoretical Framework".

CHAPTER III

THEORETICAL FRAMEWORK

I. THE CONCEPT OF INTELLIGENCE

Intelligence has no unique definition or connotation. Theorists including Spearman, Thurstone, Hebb and Guilford have all advanced reasonable explanations with empirical evidence for widely different conceptions of "intelligence".

Intelligence is primarily a "hypothetical construct" in terms of Meehl and MacCorquodale's (1948) classification. Hence it is inferred from human behavior and cannot claim direct correspondence with objective reality. The only remaining criterion on which a conception of intelligence can be justified is in pragmatic terms such as usefulness in prediction, selection and research situations.

For the purposes of this study the statistical model of intelligence described by Vernon (1950) will be adopted. This paradigm provides for a hierarchy of intellectual factors of decreasing generality. At the top of the hierarchy is the general intellectual ability factor resembling Spearman's "g". Vernon says, "The hierarchical group factor viewpoint implies that most of the variance of human abilities in daily life is attributable to g and to highly specific factors". (p. 27) Although g is not a fixed, entirely innate

characteristic it is sufficiently basic that it may be regarded as general intellectual potential. Spearman (1923) describes g as abstract reasoning ability, the ability to educe relations and correlates.

At the next level in the hierarchy the major group factors are found; these factors divide the residual variance into several major clusters particularly verbal, educational factors and practical, mechanical ones. More recently Guilford (1959) has divided the test variance for high-level airforce personnel into more than 50 group factors. The number of group factors reported is a function of the homogeniety of the group and the nature of the tests employed.

The succeeding level in the hierarchy consists of the "s" or specific factors, that is, the factors absorbing the proportion of the true variance which is unique to each test. These group and specific factors may be successively split into simpler factors, the nature and extent of which depends on the tests involved and the method of factor analysis utilized.

Pragmatic justification for the adoption of the hierarchical conception of intelligence with its concomitant emphasis on g can be approached through a discussion of "construct validity". Since this study is concerned with the construct "intelligence" and its assessment by psychometric

tests, the problem of validity, the extent to which a test measures what it is designed to measure, is obviously involved.

Traditional tests have been shown to have predictive validity by the success with which they predict to specific criteria (provided there are no dramatic changes in either the treatment or the criterion). However, when the criterion to which these tests predict are distant and general and when the treatment is adaptive, the problem of validity is much more complex. In this situation it would involve:

1. an analysis of the meaning of the test scores in terms of psychological concepts, and
2. empirical justification of the adopted approach.

First, general intellectual ability, at the top of the hierarchy, is necessarily involved in all cognitive activity. If an estimate of general intellectual functioning is required, a measure of g is imperative. Originally Spearman (1923) and a host of more recent writers including Elley (1961) demonstrated that tests with a high concentration (loading) of g require subjects to see relationships, to "go beyond the information given" and to manipulate abstract symbols, that is, they demand complex mental functioning. Selection in our schools generally involves the identification of those who can carry out these complex intel-

lectual tasks. MacArthur (1961, p.4) points out:

> Particularly for culturally-handicapped pupils in the elementary and early junior high school when there is still time to adapt instruction so as to capitalize on pupil potential an indication of a pupil's general intellectual ability, relatively independent of environmental influence, is useful in determining teaching treatments appropriate for particular pupils or groups of pupils.

To recapitulate, in order to have construct validity for children with a foreign language background in an adaptive treatment situation an intelligence test should be indicative of general intellectual ability apart from cultural and verbal influences; in terms of theory this implies high g loading with minimal loadings on other factors.

Secondly, empirical justification for this conception of intelligence is necessary because there is no absolute way of defining intelligence and no unique factor analytic solution. Hence, it can be justified primarily on the basis of "psychological meaningfulness" and usefulness.

Hunt (1961) points out that the usefulness of g and of group factors of various breadth in predicting to various criteria depend upon the age of the subject and the range of ability being considered. Spearman (1927) contended that measures of g alone suffice for such predictive tasks as grade-school placement. Cronbach and Gleser (1957), in a

discussion of communications theory and the concept of bandwidth versus fidelity make a similar observation. If the tester concentrates on collection of facts relevant to a single decision he gets a more dependable answer than if he spreads his efforts. Traditional intelligence tests in an attempt to give some indication of general intellectual ability plus some predictive validity beyond that provided by general intelligence, measure such a broad band as to serve neither function adequately. Measures of g on the other hand, although somewhat less efficient for the tasks of specific prediction, may be valuable for the selection of bright children with a cultural handicap.

Woodrow (1938) found that g is relatively immune to environmental changes. In a less general way, Elley (1961) in a study involving 271 Grade VII Edmonton children provided evidence that tests highly loaded with g, noteably the Raven's Standard Progressive Matrices, the Cattell test of g, and the Lorge-Thorndike Figure Analogies differentiated across socio-economic levels significantly less than did traditional verbal intelligence tests such as the Laycock Mental Ability Test, and are therefore relatively independent of environmental differences.

From a different point of view, Jones (1953) demonstrated that for a group of Welsh-speaking children in a

predominantly English community, the Jenkins Non-Verbal Intelligence Test, which is highly loaded with g, yielded scores comparable to those obtained by a control group of English speaking children. In other words, the bilingualism, an extreme environmental variable, has little effect on measures of g.

The most critical studies on the usefulness of tests of g are yet to be done. It is anticipated that in an adaptive treatment situation measures of g will predict such general criteria as school success more effectively than traditional tests. We therefore await some results from "adaptive treatment" experiments.

The New York Board of Education (Wrightstone, 1957) is currently involved in a long range experiment involving the effects of adaptive treatment on over 500 children from a lower socio-economic area of the city. Selection of these children was made on a series of criteria involving both intellectual and academic measures. Now that the program has been in operation for over four years, concern is being directed to selecting the intelligence tests or other variables which are able to effectively predict school success after adaptive treatment.

Preliminary reports suggest that the students undergoing this rigorous adaptive treatment have changed so much

in both intellectual and motivational aspects that no single test of intelligence has been very satisfactory in predicting adademic potential prior to the program.

In summary, the statistical construct g is involved in all complex cognitive processes, is relatively independent of environmental differences, and should be therefore, the most valuable index of intellectual potential. Tests, therefore, which are reliable measures of g have construct validity for the assessment of the intellectual ability of children with a foreign language background or other cultural handicap in an adaptive treatment situation.

II. FOREIGN LANGUAGE BACKGROUND

From the review of the related literature of both research studies and theoretical formulations presented in Chapter II it is possible to synthesize a crude theoretical framework of wide applicability for the assessed intelligence of bilinguals. Follwing this presentation, the aspects of this theory which are relevant to this study will be discussed more fully.

Since intelligence is a construct, it is defined operationally as the score on an intelligence test. These then are the major factors which modify the scores of bilingual children on intelligence tests.

1. Nature of the Tests.

   a. Tests of a Non-Verbal nature which are measures of general intellectual ability tend not to discriminate between bilinguals and unilinguals. Bilingualism appears not to interfere with mental development.

   b. Verbal intelligence tests in English discriminate between unilingual and bilingual children; the more verbal the test, the lower the score for bilingual children.

   c. Translations of a test are often as limited as the original tests themselves; the only advantage of administration in one language rather than the other is determined by the degree of bilingualism (factor 4).

2. Cultural Factors. Cultural conflict is expected to handicap bilingual children. The intensity of this conflict is largely determined by:

   a. the extent of the differences in attitudes and norms for the two cultures; the greater the differences the greater the conflict.

   b. the strength of the integrative motive and the achievement motive of the bilingual; the stronger these motives the less the conflict.

3. Socio-Economic Factors.

   a. a lower socio-economic status is related to lower verbal intelligence test scores for bilinguals as for unilinguals. Rural bilinguals tend to score lower than urban bilinguals.

4. Type and Degree of Bilingualism.

   a. Compound bilinguals tend to confuse the language system more than do coordinate bilinguals. This confusion will be felt particularly on Verbal tests.

b. The greater the degree of foreign language background, the more limited the knowledge of English and the lower the expected score on a verbal intelligence test in this second language.

c. Increased education develops increased proficiency in the second language. This factor is expected to overcome the linguistic handicap somewhat.

Since the scores for children from a foreign language background are a function of the nature of the test of intelligence, it is important that these tests be judged on some objective basis. Factor analysis provides such a basis. Elley (1961) has analized the twelve tests and subtests used in this study for a large, representative, Edmonton sample. With a knowledge of the loadings of each of these tests, it is possible to make hypothesis about the scores of bilingual children on each of these tests.

The literature suggests that for immigrants (the extreme group of bilingual children), a test employing the home language should be used for intellectual assessment. This practice is not recommended in this situation for two reasons;

1. there are thirteen languages represented in a sample of 55 bilingual cases and therefore, too few in each group to make the effort practical, and

2. there is no local reference group with which the results of these tests could be compared.

Moreover, for most of the children in this area with a foreign language handicap, English is still their best language, even if they have some familiarity with another language.

It is apparent from the literature that socio-economic factors tend to confuse the effects of bilingualism; in this study, therefore, the effects of this variable will be ruled out by using comparisons between matched groups.

Cultural factors cannot be handled directly because of measuring difficulties. It is anticipated that they will be controlled to some extent through controlling socio-economic status. Also, the Language-Background Questionnaire takes cultural factors into account somewhat. However, differential effects of various ethnic backgrounds on other aspects, such as motivation are beyond the scope of this study, and must be construed as limitations.

The final variable, that of type and degree of bilingualism is of key significance to this study. For the purpose of establishing the possible relevance of the distinction between compound and coordinate bilingualism, qualitative items regarding the age and situation in which each of the languages was acquired were included in the Language Background Questionnaire. For rating the degree of bilingualism, most of the studies in this area have used

the scale provided by Hoffman, (1935) or some abridgement of it. Because this scale was devised for use in a truly bilingual Welsh -English community and many of the questions were inappropriate a similar scale was constructed which employed items appropriate to the local situation. Since the questionnaire contains questions indicative of the degree to which the child is being assimilated into the more general Canadian culture, as opposed to confining himself to an ethnic ingroup, the score on the Language Background questionnaire can be construed as a pragmatic combination of bilingualism and biculturalism.

In summary, the intelligence test scores of children with a foreign language background will be a function of the nature of the test, the socio-economic status of the child, the age and situation in which the languages were learned, and the degree to which a language other than English is used by the child.

CHAPTER IV

DEFINITIONS, POSTULATES AND HYPOTHESES

I. DEFINITIONS

Intellectual Potential - general intellectual ability or underlying mental power, inferred from test results, which is the product of genetic influences and previous, irrevocable debilitating or stimulating environmental experiences. It represents what may be done in contra-distinction to conventional intelligence tests and achievement tests which predict what will be done in a fixed treatment situation. For the purpose of this study it will be defined operationally as the statistical factor g.

Verbal Tests of Intelligence - conventional tests requiring the reading of verbal problems. Included in this group are the California Test of Mental Maturity-Language, Laycock Mental Ability Test, Holzinger-Crowder Unifactor Battery-Series. Verbal tests are operationally defined as those possessing a significant loading on the verbal factor in Elley's (1961) factor analysis.

Non-Verbal Tests of Intelligence - those tests possessing no loading on the verbal factor. These tests require no reading for the solution of the problems. This

group includes Raven's Standard Progressive Matrices, Lorge-Thorndike Non-Verbal Intelligence Test, CTMM Non-Language, and the Holzinger-Crowder Figure Changes. This group can be broken into two subgroups, tests with loadings on g without any group factors and those with loadings on g and any other non-verbal group factor.

Foreign Language Background - the situation in which a child moves in a social circle, particularly his family, where a language other than English is used. It is a pragmatic combination of bilingualism and bi-culturalism. It is defined operationally as a score greater than ten on the Language Background Questionnaire.

Bilingual - an individual who is able to speak and/or understand more than one language in a conversational setting.

a. Compound Bilingualism - the type of bilingualism which arises when two languages are used more or less interchangeably in the same situation. For this study an individual was classified as a compound if he began to learn both languages between two and four years of age.

b. Coordinate Bilingualism - the type of bilingualism which arises when one language is used consistently in one setting and the other language is used consistently in another situation. This tends to minimize confusion of the two language systems. In this study an individual was classified as

a coordinate if he learned a language other than English in childhood but did not learn English until he was over five years of age.

II. POSTULATES

1. To provide a valid estimate of intellectual potential for Elementary School children from a foreign language background in an adaptive treatment situation, a test should meet the following criteria:

   a. possess a high factorial loading on g,
   b. possess low loading on verbal and other group or specific factors,
   c. show low relationship with language background,
   d. provide scores which are relatively stable over a period of years,
   e. consist of items which can be solved in any language and are likely to be as familiar to one social group as another, and
   f. demonstrate a satisfactory relationship to school achievement.

2. The verbal ability of bilinguals or children with a foreign language background tends to be more limited in both languages than that of unilinguals of similar age, socio-

economic status and intellectual potential. The corollary of this is that tests requiring verbal facility differentiate between unilinguals and bilinguals.

3. Tests of a Non-Verbal nature which measure abstract reasoning ability or have high factorial loading on g minimize the difference between unilingual and bilingual children of similar age and socio-economic status.

4. Foreign language background is associated with low socio-economic status. Both of these factors are related to reduced verbal intelligence test scores.

5. Type of bilingualism and degree of foreign language background are influential in determining the intelligence scores of children.

a. The greater the degree of foreign language background, the more limited the knowledge of English and the lower the expected score on a verbal intelligence test in this second language.

b. Compound bilinguals tend to confuse the language systems more than do coordinate bilinguals. This confusion will be more obvious on verbal tests of intelligence.

III HYPOTHESES

A. The selection of tests appropriate for children with a foreign language background.

1. Certain Non-Verbal tests will provide high g loadings while minimizing educational factors for a representative group of 271 Edmonton unilingual and bilingual children.

2. Certain tests will differentiate between children from a foreign language background and those from a unilingual background.

a. The mean score for a unilingual control group will be significantly higher than the mean for foreign experimental group on each of the Verbal intelligence tests.

b. The mean score for the unilingual group will not be significantly different from the mean score for the bilingual group on each of the Non-Verbal tests.

c. Verbal tests will show a significant correlation with language background.

d. Non-Verbal tests will be relatively independent of language background; correlations will not be significantly different from zero.

3. The mean scores for certain Non-Verbal tests will remain more nearly constant over the period Grades III to VII than other tests for children from a foreign language background.

4. Certain tests will show less relationship to language ability in that they require no reading and no knowledge of an intra-cultural type.

5. Verbal tests will bear a greater relationship to school achievement than do Non-Verbal tests, but among the latter, certain tests will bear a closer relationship to school achievement than others.

B. The effect of varying degrees of foreign language background on intelligence test scores.

6. Scores on Verbal tests of intelligence decreases as scores on the Language Background Questionnaire increase.

7. Scores on Non-Verbal tests of intelligence are relatively independent of the scores on the Language Background Questionnaire.

8. Compound bilingualism is a greater handicap on intelligence tests than Coordinate bilingualism.

CHAPTER V

EXPERIMENTAL DESIGN

This study was conducted as part of a larger project under the general supervision of Dr. R.S.MacArthur on the general theme "Assessing General Intellectual Ability with Minimum Cultural Bias". These studies are an analysis of selected intelligence tests in an attempt to isolate those instruments which provide a valid estimate of the intellectual potential of children from lower socio-economic strata, from homes in which a foreign language is used and from Indian and Metis communities.

In this present study, ability and achievement tests were examined in relation to the language background of Edmonton Grade VII children by means of correlations, factor analysis and comparisons of means. This investigation is parallel to that of Elley (1961) on the socio-economic bias in intelligence tests in that it employs the data and many of the conclusions of that study.

<u>The Sample</u>. The subjects for this study were drawn from three large Edmonton public schools. These schools were chosen on the basis of academic averages, socio-economic level, residential mobility, geographic location and school type, so as to be typical of the Edmonton Grade VII population.

Results from two city wide achievement surveys at the Grade III and VI level were also available for most of the subjects studied at the Grade VII level. This information was used for the testing of longitudinal hypotheses as well as for testing the representativeness of the selected Grade VII sample. Elley (1961) has demonstrated that this sample is representative of Edmonton Grade VII children by comparing the mean age, intelligence quotient, occupational level and sex of this selected sample with that of a larger, randomly chosen sample drawn from the Grade VI city wide survey. If this sample is representative of Edmonton, the results of this study can be generalized with some discretion to similar communities.

Test Instruments Used. It has already been pointed out (Postulate 1) that testing instruments must meet certain criteria if they are to be considered as valid measures of the intellectual potential of verbally handicapped children. In general instruments which qualify against this criterion and hence were included in this battery, were reliable measures of general intellectual ability which were largely non-verbal and non-numerical, and which appeared to be free from cultural bias. Moreover, they were economical, practical and convenient group tests.

For comparative purposes and to ensure a good represent-

ation of abilities in the factor analysis a few conventional intelligence tests were included. The test battery included the following tests.[1]

1. Raven's Standard Progressive Matrices (1956) has obvious face validity in that it is non-verbal, unspeeded, interesting, easily administered and not closely related to school learning. Several writers including Raven (1956) and Vernon (1950) have reported g loadings of over .79. The test has a reported re-test reliability of .88. The test is composed of sixty items classified in five sets, each of which has a particular principle for solution. The first problem of each set is simple enough to be self-explanatory with successive items becoming increasingly difficult.

Each item consists of a matrix of nine geometric figures, one of which is omitted and must be selected from among six or eight alternatives. Generally, the items demand that the subject educe the principles operating in the matrix and go on to educe the nature of the missing element, in other words, to "go beyond the information given". Hence the test is heavily loaded with g with no significant group factors loadings.

2. IPAT Cattell Test of g, "Culture-Free", Scale Two, was designed as a culture-free test of g factor. The scale

---

[1] This description of the tests and the reasons for inclusion is taken largely from Elley (1961) p. 72-85.

consists of four sub-tests, with a total of forty-six items of a perceptual nature. It is economical and easy to administer. Cattell (1958) reported immediate re-test reliability of .85.

The first test, contains twelve items involving series. The first three figures change progressively according to some rule, and the subject is required to select the fourth from among five alternatives. The second test consists of fourteen items, each containing five different geometrical designs. The subject is required to select the design that does not belong with the others. The third subtest is similar in construction to the Raven's and the forth involves the identification of a design with the same spatial arrangement as the stimulus design.

This test has the disadvantage of requiring considerable instructions to solve the first items.

3. The Lorge-Thorndike Non-Verbal Intelligence Test, Level 4, (1957), another non-verbal, unspeeded, economical test has a reported split-half reliability coefficient of .93 and an alternate forms reliability of .78.

The three sub-tests require more verbal instructions than the Raven's, and like the preceeding tests it relies only slightly on verbal skills and school learning.

The first subtest requires the subject to select from

among five alternatives the design that is similar in principle to the three stimulus designs. Then follows the number series subtest which is obviously loaded on the number factor. The third subtest consists of figure analogies in which the subject must educe a relationship between the first two designs, and then apply this relationship to a new situation. This test is regarded as one of the best on the market today (Buros, 1958).

4. Holzinger-Crowder Uni-Factor Test (1952), a multi-aptitude test, from which three subtests were selected for inclusion in this battery, is a well constructed and well standardized test. The non-verbal reasoning subtests selected for this battery showed split-half reliability coefficients of .93 on Grade VII children. One subtest, Figure Changes, appears to be less dependent on past learning as it uses only perceptual items requiring ability to see and use relationships. The Mixed Series test which consists of number and letter series completions is expected to have loadings on factors other than g. The other subtest is a Spatial test requiring the subjects to decide whether two boots were viewed from the same or different sides.

5. California Test of Mental Maturity (CTMM), Short-Form, Elementary (1957) is a popular, conventional intelligence test, and hence was included for purposes of comparison with

the Non-Verbal, "culture-free" tests. This test furnishes separate factor scores for spatial, logical, numerical, and verbal abilities, a language, a non-language, and a total intelligence quotient. The reported split-half reliability coefficients for the subtests range from .87 to .93. The validity and usefulness of the separate factor scores can be criticized in that they were set up on logical grounds instead of an empirical factor analytic basis. The test as a whole is practical and convenient.

The spatial factor is made up of two tests, one consisting of pictures of separate hands or feet in various positions which are to be labelled as left or right, the other of a figure which has to be recognized in rotated form among a group of five. The first logical subtest is a classification test requiring the pupil to select a picture which conforms to the principle common to a group of familiar objects, the other consists of a list of logical syllogisms in verbal form. A number series and an arithmetic reasoning test make up the numerical factor, and a vocabulary multiple-choice test yields a verbal concepts factor.

6. Laycock Mental Ability Test (1933) is another widely used conventional intelligence especially included for the purposes of comparison. It is based on Spearman's theory of intelligence and so constructed to measure the ability to

educe relations and correlates through the use of verbal, numerical and abstract symbols. The test reportedly correlates with other popular intelligence tests about .8, and has a re-test reliability of .80.

7. California Achievement Battery Junior High (1957) is parallel to that given to the sample studied at the Grade III and VI levels and was included for factor analytic purposes.

8. Colored Progressive Matrices (1947) was used in the Grade III survey and because of the similarity to the 1956 edition used in Grade VII, it was used for comparing changes in intelligence score over a four year period.

9. California Test of Mental Maturity, Primary (1953) was another instrument employed in the Grade III survey. It is similar to the CTMM Elementary and for this reason it too was used for comparisons over the four year period.

10. Occupational Scale is an objective scale for the measurement of socio-economic status compiled by Blishen from Canadian census data on the basis of average income and years of training. Information on parental occupation was collected by Elley from the Grade VII students by means of a questionnaire.

11. Home Index Questionnaire was prepared by Elley (1961) for purposes of validating the Occupational Scale. A

single additional question inquired into the degree to which a foreign language was spoken in the home.

12. Language-Background Questionnaire somewhat similar to that devised by Hoffmann (1935) was prepared specifically for this study. The original scale was designed for use in Welsh-English bilingual communities and consisted of about fifty items. After careful consideration it was decided to include only sixteen items which were closely associated with the actual use of a language other than English in the home. Each item has four alternatives of increasing degree with these alternatives receiving 0, 1, 2, 3 points respectively. The final questionnaire (See Appendix A) was employed in two pilot runs in Calgary and Faust, on the basis of which the administrative instructions were modified somewhat. Since the units of the questionnaire are not truly interval in nature the scores were grouped into four crude categories. By correlating the results of this questionnaire with that of the item from the Home Index Questionnaire, a Contingency coefficient of .74 (maximum less than .89) was established. This relationship may be construed, somewhat tenuously, as a validity coefficient if the results of the Home-Language Questionnaire can be regarded as an independent objective criterion. Since no subtleties and no hypothetical constructs are involved here, the validity is relatively obvious.

<u>Data Gathering Procedures</u>. Data on the same set of individuals were gathered over a four year period.

In May 1956, more than 3,500 Edmonton public school Grade III children were given the following tests:

1. California Short Form Test of Mental Maturity (Primary, 1953).

2. Raven's Coloured Progressive Matrices (1947).

3. California Achievement Battery (Primary, 1951).

4. Gates Reading Tests.

In May 1959, the same children, then in Grade VI were tested with a parallel battery of tests including:

1. California Short Form Test of Mental Maturity (Elementary, 1957).

2. California Achievement Battery, (Junior High, 1957).

In May 1960, a sample of 430 children, then in Grade VII were given the following tests:

1. California Short Form Test of Mental Maturity (Elementary, 1957).

2. Raven's Progressive Matrices (1956).

3. IPAT Cattell Test of g - Culture-Free, Scale 2.

4. Lorge-Thorndike Non-Verbal Intelligence Test, Level 4.

5. Holzinger-Crowder Uni-Factor Tests - Spatial (Number 3) Reasoning (Numbers 7 and 8).

6. Home Index and Occupational Questionnaire.

All this testing was done under the direction of University of Alberta personnel, the 1960 battery being closely supervised by Elley (1961).

In June 1961, the 430 subjects then in Grade VIII were traced and requested to complete the following form:

1. Language-Background Questionnaire.

Since the purpose of this study revolves around this variable, it was essential to ensure maximum uniformity in the collection of these data. To this end, the writer personally administered the questionnaire to each class. Since several individuals had been changed from the general group with which they had been associated in Grade VII it was necessary to follow up many individual cases. This questionnaire was administered to 350 of the 430 individual subjects. More important is the fact that of the seventy students who in the Home Language Questionnaire reported at least some use of a language other than English, only two could not be traced.

<u>Analysis of Results</u>. In the Grade VII survey, all children in the three schools were tested; however, Elley (1961) included only 261 of these pupils in the factor analysis. Elley has shown this group to be representative of Edmonton Grade VII children.[1] In the present analysis, all

---

[1]For a more comprehensive analysis of the representativeness of this sample see Elley (1961) p. 95 - 102.

430 children in the original sample were considered as potential subjects.

The results from the administration of the Language-Background Questionnaire were used as a basis for the formation of three groups of children differing in the extent to which a foreign language was used in the home. The first of these groups having no record of any language other than English in the home was the group from which the unilingual or control group was chosen. The other extreme group who scored highest on the Language-Background Questionnaire were called the experimental or bilingual group. An intervening group of individuals with a moderate degree of foreign language background was included for comparative purposes.

For the purposes of this study socio-economic status was taken as the score on the Blishen Occupational Index as Elley (1961) has demonstrated that for this group it was an effective socio-economic scale. To eliminate socio-economic factors these three groups were matched on socio-economic status for mean and standard deviation.

To facilitate the comparisons among tests and subjects, all intelligence tests scores were converted to T-scores on the basis of the larger Edmonton Grade VII sample. The specific hypotheses were tested by a variety of methods.

Hypothesis I requires the resulting information from a Principle Components factor analysis carried out by Elley to select the tests of intelligence with the specified factorial characteristics.

Hypothesis II a and b involved a t-test for the significance of the difference between the means of the experimental and control groups on each test of intelligence.(Walker and Lev., 1953, p.155). Parts c and d were tested by a point-biserial r. Dichotomized Language Background scores were correlated successively with the scores on each test of intelligence (Garrett, 1959, p. 380). The resulting r's were tested for significance.

Hypothesis III was tested using a Non-Verbal test that has met the criteria specified in Hypotheses I, II, IV and V, and a representative Verbal Test. Differences between the corresponding test scores were found for each individual; the means and variances of these differences were then tested for significance.

Hypothesis IV was examined by inspection of both the content and administration for each of the tests.

Elley (1961) found the median correlation between the average school mark and each intelligence test for the thirteen

classes involved. The results of this procedure were used to test Hypothesis V.

When all of these hypotheses had been examined each test and subtest was examined in the light of the findings in an attempt to isolate those instruments which possess construct validity for the testing of children from a foreign language background in a potentially adaptive treatment situation.

Hypotheses VI and VII were examined by comparing the means of the three groups previously set up. Significance of the difference between the extreme groups was tested in hypothesis II. The means of the middle group were indicative of the nature and consistency of the trend.

Hypothesis VIII was tested using the results of the qualitative aspects of the Language Background, including the age at which each language had been learned and where it had been learned. A similar dichotomy was set up between immigrants and second-generation Canadians. These groups turned out to be similar in constitution to the original groups set up on the basis of Language-Background score and hence little additional testing of differences was required. Conclusions and implications were then drawn with respect to the effect of varying degrees of foreign language background on intelligence tests.

CHAPTER VI

PRELIMINARY RESULTS

The Language Background Questionnaire

Since this questionnaire was not identical with any existing scale, it was necessary to investigate its effectiveness as a measure of degree of foreign language background. First, it was necessary to break up the scores into ordinal categories. A graphical representation of the frequencies of each score suggested three major divisions. (Figure 1). On the basis of these divisions, the following categories were formed:

Descriptive Classification	Lang.-Back. Score	Category	Number of Individuals
Never	0-3	1	340
Seldom	4-9	2	39
Quite often	10-23	3	33
Most of the time	24-	4	22

One week after the original administration, the Language Background Questionnaire was re-administered to two classes with a combined attendance of sixty-three for the purposes of calculating re-test reliability. Using these categories as an interval scale the two administrations were correlated using Pearson's Product Moment Correlation.

FIGURE 1

FREQUENCY POLYGON OF LANGUAGE BACKGROUND SCORES DIVIDED INTO CATEGORIES FOR NINETY CHILDREN WHO INDICATED SOME DEGREE OF FOREIGN LANGUAGE BACKGROUND

This produced a re-test reliability coefficient of .98. To examine the internal consistency of the questionnaire an analysis of the frequencies for each alternative of each item was made for each category. This revealed that the score on each item was closely related to that on each other item and to the test as a whole. That is, as total scores on the questionnaire increased all items showed a progressive increase in the degree of bilingualism implied by the answer to the question.

The question of validity is not as easily handled. The questionnaire has obvious face validity in that it asks in a straightforward manner for the information required. If it can be assumed that denial or lying was absent, the responses should be valid.

Emperical validation was sought by correlating the results from the question on Elley's Home Index Questionnaire and the results from the Language Background Questionnaire. For the 130 subjects in the experimental and control groups a Pearson Product Moment Correlation Coefficient of .79 was found between these two instruments. Since the information required on these two forms is not strictly parallel, the correlation gives some evidence of validity. The limitation of this interpretation lies in the fact that both scales employ subjective ratings by the same individuals.

Another major problem faced in using the results of the Language Background Questionnaire was the establishment of a cutting point, that is, a point below which scores would not be considered as implying a home in which a foreign language was used. Originally, it was thought that some subjects may occasionally speak to playmates in French (it being a subject of instruction in school) and, hence, receive one point on the questionnaire. To exclude such contingencies it was decided to place this cutting point at a score of four.

However, after a series of analyses of the group nearest but above the cutting point, it became apparent that Language Background was having no influence on this group. The degree of foreigness was so mild that it was irrelevent. Hence, the cutting point was raised to ten. An analysis of the distribution (Figure 1) shows the appropriatness of such a decision, in that a score of nine is the first apparent break in the distribution of the Language Background scores.

The Comparison Groups.

On the basis of Language Background categories three groups, a control, a moderately bilingual and a bilingual group, matched on socio-economic status were set up for the testing of the hypotheses.

In the first attempt to select a control group it was thought to match the groups for age. This introduced a com-

plicating factor. The age of the bilingual group was found to be higher than that of average Grade VII children; it was assumed that this retardation was due to the influence of language difficulties. In selecting a control group equal in age but without foreign language background overaged children were, therefore, required. However, these overaged children are not retarded because of a language problem but presumably because they have less scholastic aptitude. Hence, in comparing these groups, bilingual children would appear to do better than unilingual ones. The correlations between age and Raven's Matrices (Table II) for each of these groups supports this contention.

To overcome this distorting influence the control group was reselected. This time age was sampled randomly but socio-economic status was selected so that the mean was equal to that of the other groups. The age differential among these groups will be relevant in comparisons of actual scores across groups and will tend to minimize these differences. However, since this age difference between groups is the same for all tests, comparison among tests are in no way hampered. The significancesof the differences among five relevant variables for three final experimental groups are shown in Table II.

An examination of this table shows that the three groups differ primarily in the degree of foreign language background. Group 1, the unilingual, control group consists of

TABLE II

SIGNIFICANCE OF THE DIFFERENCES FOR FIVE RELEVANT VARIABLES ACROSS THE THREE GROUPS USED IN THE TESTING OF HYPOTHESES

		Group I Uniling.	Group II	Group III Bilingual	Obtained t,z,or F	Level of Sign.
N		40	33	22		
Lang.-Back.		0	2	3	F = 20.17	.01
Soc.-Ec. Status	Mean	47.20	47.47	45.64	F = 1.35	N.S.
	S.D.	4.93	5.53	5.79	1.09<F<1.38*	N.S.
	M	18	12	10		
Sex	F	22	21	12		
	% M	.46	.36	.45	.08<t<.83*	N.S.
Age	Mean	13-0	13-3	13-7	F = 3.00	.05
r(Age x Raven)		-.23	-.14	+.24	1.33<z<1.68*	N.S.

*More than one test was required, hence the maximum and minimum values for F, z or t were recorded.

forty subjects who fell into category 1, that is, they reported no language other than English in the home.

Group II consists of subjects in language background category 3; these subjects have a moderate degree of familiarity with a foreign language. English is the major language used by this group but most of the subjects understand quite well and speak part of the time in a language other than English. Most of the subjects in this group were born in Canada and learned to speak English and another language concurrently during childhood. One-third of this group speak Ukranian or Polish and about one-quarter speak German. A few of the subjects use Dutch, French or Chinese.

Group III, the bilingual group consists largely of subjects who, although they speak English, have some other language as their primary language. These subjects all fall into language category 4. Two-fifths of this group have lived in Canada six years or less; only one-quarter of the group was born in Canada. Three-quarters of the group were classified as immigrants. One-half of the subjects in this group speak German, with the remainder of the subjects speaking Ukranian, Dutch, Yiddish, or Norwegian. The mean age of this group is over one-half year above that of the control group, a factor likely attributable to reclassification upon entry to Canadian schools or to retardation due to a language handicap.

Summary

Preliminary examiniation of the data revealed these important conclusions.

1. The Language Background Questionnaire is a consistent if somewhat crude instrument capable of yielding four categories of increasing degree of foreign language background. The re-test reliability was .98 and the correlation of this test with an independent measure was .79 for this sample of Grade VII children.

2. Children scoring less than ten on the Language Background Questionnaire reflected such a slight degree of foreign language background as to be negligable and hence were excluded from the major analysis. The implication of this observation in that our third and forth generation Canadians who still bear a trace of a language other than English are generally so well integrated as to be little different in verbal ability from their unilingual neighbors.

3. The control (Group I) chosen randomly except for the restriction of socio-economic status and language background was found to be significantly younger than the other two groups. This age factor is expected to minimize differences between groups and hence yield results which minimize Type I errors but are prone to Type II errors. Comparisons between tests are not expected to be seriously hampered.

4. These procedures provided three groups matched on socio-economic status which can be used to assess the influence of bilingualism on intelligence test scores.

CHAPTER VII

THE FACTOR ANALYSIS

In order to examine the hypothesis that tests having construct validity, for the assessment of the intellectual ability of children from a foreign language background, should have a high loadings on g without group factors, two factor analyses of the tests and subtests in the battery were conducted. This procedure was carried out by Elley (1961) for this same sample, hence, the results of this analysis will simply be adopted for use here. The treatment in this chapter will be paraphrased from Elley.

The purpose of the elaborate statistical technique known as "factor analysis" is to determine the relationships of tests with each other in order to find the dimensions common to tests in the battery and the extent to which each test is loaded on each of these dimensions. Harman (1961, p.5) points out: "In all cases, ..., factor analysis does give a simple interpretation of a given body of data and thus affords a fundamental description of the particular set of variables analyzed."

The model used in factoring was Hotelling's Principle Components which is particularly adapted to the use of electronic computers. This solution provides for orthogonal factors

of decreasing magnitude each of which accounts for a maximum proportion of the communality. This program was written and carried out by Elley on the L.G.P. 30 computer using squared multiple correlations as the original estimate of communality.

Following the original analysis a series of rotations were made to the three criteria of positive manifold, maximum g and psychological meaningfulness. This procedure was applied to two batteries of tests with the following results.

The Sub-Test Analysis

The rotated factor pattern for twenty-one variables used in the first analysis is shown in Table III. The interpretation of these factors is fairly straightforward.

Factor I, having high positive loadings on nearly all variables and accounting for over fifty percent of the common variance is the general factor, g. The three highest loadings on this factor are provided by Raven's Matrices, Cattell's Test of g, and the Figure-Analogies subtest of the Lorge-Thorndike.

Factor II with its high loadings on two vocabulary tests and arithmetic tests, Elley labels as the v:ed (verbal-educational) factor in line with Vernon's (1950) analysis.

Factor III is a socio-economic factor with loadings exclusively on socio-economic variables.

Factor IV with its loadings on arithmetic and series tests, Elley labels as n (numerical).

TABLE III

ROTATED FACTOR SOLUTION AND COMMUNALITIES FOR TWENTY-ONE VARIABLES.[1]

(Loadings below .20 omitted)

	I_3	II_2	III_1	IV_1	V_1	h^2
Occupational Status	.251	-	.683	-	-	.583
Home Index	.247	.205	.669	-	-	.602
Raven Matrices	.708	-	-	-	-	.542
Cattell g	.754	-	-	-	222	.626
Lorge-Th. Fig. Class	.582	-	-	-	-	.355
Lorge.Th. No. Series	.547	-	-	395	-	.486
Lorge. Th. Fig. Anal.	.738	-	-	-	-	.553
CTMM Number	.639	.201	-	.331	-	.563
CTMM Verbal	.468	.658	-	-	-	.694
CTMM Spatial	.612	-	-	-	-	.428
CTMM Logical	.660	-	-	-	-	.476
Laycock	.676	.509	-	-	-	.735
Holz-Crowder Spatial	.397	-	-	-	.396	.326
Holz-Crowder Series	.457	.215	-	.453	.303	.554
Holz-Crowder Fig.Changes	.498	-	-	.211	.369	.440
Calif. Rdg. Vocab.	.335	.743	-	-	.264	.747
Calif. Rdg. Comp.	.501	.619	-	-	-	.681
Calif. Arith. Reas.	.460	.341	-	-	-	.369
Calif. Arith. Fund.	.450	.436	-	.442	-	.629
Calif. Language	.417	.589	-	-	.299	.641
Calif. Spelling	.201	.618	-	-	-	.443
% of Common Variance	51.18%	25.23%	9.11%	7.80%	6.69%	100%
% of Total Variance	27.94%	13.78%	4.97%	4.26%	3.65%	54.60%

[1] From Elley, 1961, p. 117.

Factor V is labelled somewhat tentatively by Elley as a speed factor.

Analysis of Total Tests

The rotated factor solution for the eleven total tests in the battery is shown in Table IV. The intercorrelations among tests were accounted for by the three major factors in the first analysis, that is, a large g factor, a v:ed factor and a number factor. The proportion of the common variance accounted for by each successive factor is also shown in Table IV.

Subtests employed in both analyses show the stability of the factors for this sample in that the mean discrepancy in corresponding loadings is less than .06.

Summary

The factor analytic results provide the solutions for Hypothesis I. It is apparent that certain Non-Verbal tests of intelligence provide measures of the one broad intellectual ability factor while having negligible loadings on group factors. The tests most highly loaded with g and negligible loadings on other factors are the Raven Matrices, Cattell Test of g and the Figure-Analogies sub-test of the Lorge-Thorndike Test. These tests measure general intellectual ability apart from verbal, numerical or other educational skills.

A second cluster of tests including the conventional measures such as the CTMM Language Test, the Laycock, the

TABLE IV

ROTATED FACTOR SOLUTION AND COMMUNALITIES
FOR ELEVEN TOTAL TESTS[1]

(Loadings below .20 omitted)

Variables	I_2	II_1	III_1	h^2
Raven Matrices	.780	-	-	.634
Cattell g	.792	-	-	.639
Lorge-Thorndike	.740	-	.325	.678
CTMM Language	.580	.616	-	.720
CTMM Non-Lang.	.624	-	.349	.516
Laycock	.634	.573	-	.730
Holz-Crowder Series	.399	.338	.531	.556
Holz-Crowder Fig.Changes	.521	-	.256	.374
California Reading	.444	.784	-	.819
California Arithmatic	.430	.509	.343	.562
California English	.361	.702	-	.654
% of Common Variance	55.97%	32.83%	11.20%	100.00%
% of Total Variance	35.02%	20.55%	7.01%	62.07%

[1]From Elley, 1961, p. 121.

Holzinger-Crowder Series and the California Achievement Tests have high v:ed loadings and, therefore, demand familiarity with the English culture, particularly the English Language.

The tests in the third group are those with significant number factor loadings, including the CTMM Numerical and the Series subtests from the Lorge-Thorndike and the Holzinger-Crowder Tests. These tests too are presumably "culture-reduced" since they have no predictable bias for or against bilingual children, numbers being a more universal system than the English language. However, even if they are culture-fair, they cannot be construed as highly valid measures of the intellectual potential of any group of individuals because of their greater dependence on acquired skills and their generally smaller g loadings.

The use of these factor analytic results in describing what each test measures is limited somewhat in that the tests would have slightly different loadings if only subjects with a foreign language background were included. This expected difference is not so great as to make the present analysis irrelevent; it does, however, show that it must be used with some caution.

CHAPTER VIII

THE EFFECT OF LANGUAGE BACKGROUND ON INTELLIGENCE TEST SCORES

Hypothesis II suggests that certain Non-Verbal tests will differentiate less between unilingual and bilingual children than will conventional instruments. To test this hypothesis Group I and Group III were taken to represent unilingual and bilingual children respectively. Group II was so similar to Group I that it was not classified as bilingual and hence was omitted from this analysis.

Two statistical approaches were taken in testing this hypothesis. First the mean T-scores on each test of intelligence was compared for the two groups, and secondly, correlations between each test and language background scores for these groups were calculated.

Comparison of Means

To facilitate the comparisons between tests, all intelligence test scores were converted to T-scores based on the entire Grade VII sample. The means were then tested for significant differences. (Walker and Lev, 1953, p. 155). The use of T-scores gave assurance that the scores were normally distributed; the homogeniety of variance was tested by the use of an F test (Walker and Lev,1953,p.192).These measures

provided evidence that the data met the assumptions underlying the use of the t-test. In order to evaluate the obtained t's the region of significance was set arbitrarily at $P \leq .05$; the region for high significance was set at $P \leq .01$.

Hypotheses II a and b were stated statistically:

for Hypothesis IIa $H_1: \mu_I > \mu_{III}$ for Verbal tests

for Hypothesis II$_b$ $H_o: \mu_I \neq \mu_{III}$ for Non-Verbal tests

where, μ_I - mean of unilingual group, Group I

μ_{III} - mean of bilingual group, Group III

H_1 was examined using a one-tailed test of significance.

Ho was examined using a two-tailed test of significance.

Thus thirteen Non-Verbal and five conventional Verbal tests were examined across the unilingual and bilingual groups and t values were tested and tabulated in Table V.

This table clearly shows that the Non-Verbal tests as a group (tests with no significant loading on the v:ed factor), minimize the bias against bilingual children and in some cases appear biased in their favor.

The tests which showed significant differences between the groups, that is, the CTMM and the Laycock, are highly verbal, conventional tests. The CTMM Total only approached significance because of the opposing influence of the CTMM Numerical subtest. This differentiation between bilingual and

TABLE V

SIGNIFICANCE OF THE DIFFERENCE OF T-SCORE MEANS FOR UNILINGUAL AND BILINGUAL GROUPS ON SELECTED INTELLIGENCE TESTS GROUPED ON THE BASIS OF FACTOR LOADINGS

Factor Loadings	Test	Group I Uniling.	Group III Bilingual	Difference I - III	t	Level of Sign.
g + n	Holz-Crowder Fig. Changes	46.84	50.95	- 4.11	-1.47	N.S.
	CTMM Numerical	46.63	49.57	- 2.94	-1.45	N.S.
	Holz.-Crowder Series	47.00	49.74	- 2.74	-1.01	N.S.
	Lorge.-Thorn. Number Series	48.97	50.71	- 1.74	-0.75	N.S.
	CTMM Non-Lang.	47.13	48.52	- 1.39	-0.52	N.S.
g	CTMM Spatial	47.18	47.71	- .53	-0.20	N.S.
	Raven Matrices	48.68	48.45	0.23	0.08	N.S.
	Cattell Test of g	47.79	47.05	0.74	0.28	N.S.
	Lorge-Thorn. Fig. Class.	49.10	48.19	0.91	0.39	N.S.
	Lorge-Thorn. Total	47.36	46.05	1.31	0.54	N.S.
	Holz.-Crowder Spatial	49.73	46.50	3.32	0.85	N.S.
	Lorge-Thorn. Fig. Analogies	46.44	43.57	2.87	1.17	N.S.
g + v:ed	CTMM Total	47.16	44.14	3.02	1.20	N.S.*
	CTMM Verbal	48.63	43.38	5.25	1.97	.05*
	CTMM Language	48.11	42.76	5.35	2.00	.05*
	Laycock	49.03	42.90	6.13	2.52	.01*
	CTMM Logical	49.63	42.71	6.92	2.76	.01*

*1 tailed test

unilingual groups, plus the fact that tests have high loadings on the v:ed factor clearly shows that these verbal tests have a minimizing effect on the scores of children with a foreign language background.

Of the remaining Non-Verbal tests, the group of tests giving the highest scores to the bilingual groups are those involving numerical ability, that is they all have high loadings on the number factor. An illconsidered conclusion would be that tests of numerical ability are the best estimates of the intelligence of bilingual children. However, in this case the highest score is not necessarily the best estimate of intellectual ability. It is more likely that these bilingual children have superior numerical ability not superior general aptitude. This conclusion seems reasonable in that the bilingual group is approximately .7 years older than the control group. Since arabic numbers are taught in most schools regardless of the language of instruction, it is reasonable that this extra half year of schooling should increase the facility of these children with numbers. This schooling would have direct impact on numerical ability but would not be expected to have such a direct bearing on general abstract reasoning ability or g. As these test scores all reflect the extra schooling they are not as valid for estimating intellectual potential as tests which are independent

of numerical ability.

The remaining tests include those which show very little difference between the unilingual and bilingual groups. The most promising of these are the Raven Matrices, Cattell Test of g, Lorge-Thorndike Figure Classification and the CTMM Spatial all of which show negligible differences between groups, negligible loadings on group factors and large loadings on g. The CTMM Non-Language and the Lorge-Thorndike Total appear to be fairly adequate. Another important factor about this group of tests is the fact that they show very little variation in mean from one to another. This mean score is near that expected on the basis of the lower socio-economic status of these groups.

To show the effect of a mild degree of bilingualism, Group II was compared with the two major groups. The mean T-scores for the three groups on each intelligence test are presented in Table VI. Although the differences between Group I and Group II were too small to be significant, they nearly all lie in the direction expected on the basis of the foreign language background. This confirms the consistent trend noticed in the preceeding analysis. It also shows that for a moderate degree of foreign language background the bias of conventional tests is small.

TABLE VI

MEAN T-SCORES ON SELECTED INTELLIGENCE TESTS FOR THREE GROUPS FORMED ON THE BASIS OF LANGUAGE BACKGROUND SCORES

Factor Loadings	Test	Group I Uniling.	Group II	Group III Bilingual
g + n	Holz.-Crowder Figure Changes	46.84	47.31	50.95
	CTMM Numerical	46.63	46.77	49.57
	Holzinger-Crowder Series	47.00	48.76	49.74
	Lorge-Thorndike Number Series	48.97	47.33	50.71
	CTMM Non-Language	4 7.13	49.29	48.52
g	CTMM Spatial	47.18	50.45	47.71
	Raven Matrices	48.68	49.40	48.45
	Cattell Test of g	47.79	47.42	47.05
	Lorge-Thorndike Figure Class.	49.10	48.20	48.19
	Lorge-Thorndike Total	47.36	46.60	46.05
	Holzinger-Crowder Spatial	49.73	47.31	46.50
	Lorge-Thorndike Figure Analogies	46.44	46.77	43.57
g + v:ed	CTMM Total	47.16	47.48	44.14
	CTMM Language	48.11	47.39	42.76
	CTMM Verbal	48.63	48.16	43.38
	Laycock	49.03	45.53	42.90
	CTMM Logical	49.63	47.38	42.71

Correlations with Language Background

In order to test Hypothesis II c, and d, that is, that Verbal tests of intelligence will show a significant correlation with Language Background while Non-Verbal tests will not, the Language Background category of the two groups, unilingual and bilingual, was correlated successively with each intelligence test using the Point-Biserial method. (Garrett, 1958, p. 380) The statistical hypotheses can be stated:

for Hypothesis II_c H_1: $r_{V\text{-}LB} > 0$

for Hypothesis II_d H_0: $r_{(N\text{-}V\text{-}LB)} = 0$

where,

$r_{V\text{-}LB}$ is the correlation of verbal intelligence and language background, and,

$r_{N\text{-}V\text{-}LB}$ is the correlation of Non-Verbal intelligence and language background.

H_1 was examined using a one-tailed test of significance;

H_0 was examined using a two-tailed test of significance.

The Point-Biserial r's and their level of significance are reported in Table VII.

This Table shows results parallel to those found by comparison of means in Table V. The same Verbal tests, the CTMM Logical, CTMM Verbal, CTMM Language and the Laycock showed significant correlations with language background.

Although none of the other correlations is significant, the same trend occurs as in the comparison of means. The

TABLE VII

POINT -BISERIAL CORRELATIONS OF INTELLIGENCE TESTS WITH LANGUAGE BACKGROUND FOR GROUPS I AND III

Test	r with Language Background	Level of Significance
Holzinger-Crowder Figure Changes	-.20	N.S.
CTMM Numerical	-.19	"
Holzinger-Crowder Series	-.14	"
Lorge-Thorndike Number Series	-.10	"
CTMM Spatial	-.03	"
CTMM Non-Language	-.07	"
Raven Matrices	-.02	"
Cattell Test of g	+.04	"
Lorge-Thorndike Figure Class.	+.05	"
Lorge-Thorndike Total	+.07	"
Holzinger-Crowder Spatial	+.13	"
Lorge-Thorndike Figure Analogies	+.15	"
CTMM Total	+.16	N.S.
CTMM Verbal	+.26	.05
CTMM Language	+.26	.05
Laycock	+.33	.01
CTMM Logical	+.34	.01

One-tailed tests of significance

tests again showing least relationship with language background were the Raven Matrices, Cattell Test of g, Lorge-Thorndike Figure Classification and the CTMM Spatial. Since Hypothesis II required tests with little relationship to language background, these tests again meet the criteria.

The third group of tests showing a negative relationship with language background again includes all those tests with loadings on the number factor. Since these tests are not independent of Language Background scores they do not meet the criteria required in the selection of maximally valid instruments.

Summary. On the basis of comparisons between intelligence test means of Groups I and III, the unilinguals and the bilinguals, matched on socio-economic status, and by finding the correlations of Language Background with each test for these groups, three conclusions were reached.

1. Certain Non-Verbal tests including Raven Matrices, Cattell Test of g, Lorge-Thorndike Figure Classification and the CTMM Spatial were found to be independent of the language background of bilingual children. Other tests showing promise were the Lorge-Thorndike Total and the CTMM Non-Language.

2. Children from a foreign language background score significantly lower on Verbal tests of intelligence than do

unilinguals.

3. Non-Verbal tests with loadings on the number factor appear related to language background and perhaps bias in favor of the bilingual group.

Besides these conclusions it also became apparent that children with a milder degree of foreign language background (Group II) were not significantly different from the unilingual group of the same socio-economic background, on either Verbal or Non-Verbal tests. It may be that their language handicap is so slight as to have little adverse effect on their performance on intelligence tests.

CHAPTER IX

ADDITIONAL FINDINGS

Changes in Intelligence over a Four Year Period

Since certain Non-Verbal tests are to be considered as measures of intellectual potential, it was suggested by Hypothesis III that these tests produce estimates of intellectual ability that are relatively constant over a four year period. Verbal tests are expected to yield an underestimate of intellectual potential for bilingual children in an adaptive treatment situation. Since adaptive treatment was only slightly involved, the direction of the changes over the four year period could not be specified; it remained only to consider the relative instability of these test scores for children from a foreign language background.

The testing of this hypothesis included only the Raven Matrices and the California Test of Mental Maturity because these were the only tests administered at both Grade III and Grade VII.

A limitation in this approach is the possibility that the tests administered at the two grade levels are measuring different abilities. This possibility is minimized by the fact that the Raven Colored Matrices administered at the Grade III level has been shown to have factorial dimensions similar to that possessed by the Raven Standard Matrices used in Grade

VII (MacArthur, 1960). Moreover, comparison of the items in the two forms for both the Raven Matrices and the CTMM Language reveals general similarity of content in the comarable forms.

The sample used for the testing of this hypothesis was smaller than the samples used in the other parts of this study. Of the twenty-two in the bilingual group (Group III) only eleven were present for Grade III and only seventeen of the thirty-three in Group II were available. Since only one-half of the original sample was present on both occasions, the results from the testing of this hypothesis must be inconclusive. In order to find if this small group was representative of the larger group from which it was drawn, comparison of the sample and subsample was made on several basis; the results are reported in Table VIII.

An examination of this table shows that the selected subsamples IIa and IIIa do not exactly represent the larger samples from which they were drawn although no large differences were observed. In these selected sub-samples Group IIIa has a higher mean score on the CTMM Language, while both groups have slightly higher mean ages, than the larger samples from which they were drawn. To increase the size of the N and to make the statistics more stable the sub-samples Group IIa and IIIa were combined for the testing of this hypothesis.

TABLE VIII

COMPARISONS OF TWO SUB-SAMPLES IIa AND IIIa WITH THE PARENT SAMPLES II AND III FROM WHICH THEY WERE DRAWN

Group	N	Years in Canada	Age	Soc-Ec. Status	Raven Matrices	CTMM Language
II	33	11.0	13.3	47.20	49.40	47.39
IIa	17	11.8	13.1	47.18	50.56	47.00
III	22	7.9	13.7	45.64	48.45	42.76
IIIa	11	8.1	13.4	46.18	48.18	46.55

The hypothesis was tested by using difference scores representing the difference between the Grade III T-score and the Grade VII T-score for each subject on first the Raven's and then the CTMM. The means and the variances were then examined to find the magnitude and direction of changes over the four years. The gains in T-scores and the variances of these gains are shown in Table IX. .

From this table it is apparent that the mean Raven's score showed a slight decrease over the four years while the the CTMM Language mean remained constant. The difference between tests, however, was not significant. More important is the fact that the variance of the differences between the Grade III and Grade VII scores for the Raven's was significantly smaller than the corresponding variance for the CTMM Language. In other words, the Raven's scores tended to remain fairly constant while the CTMM scores tended to either increase or decrease over the four year period. On the basis of this comparison it may be concluded tentatively that the Raven's, a Non-Verbal test highly loaded with g provides a more stable extimate of the general intellectual ability of bilingual children than does the CTMM Language.

<u>The Problem of Face Validity</u>

In order to possess high face validity as measures of general intellectual ability of children with a foreign language background tests are expected to have certain obvious

TABLE IX

MEAN AND VARIANCE OF GAINS IN T-SCORES FROM GRADE III TO VII FOR BILINGUAL CHILDREN ON TWO INTELLIGENCE TESTS

	Raven Matrices	CTMM Lang.	Obtained t or F	Sign.of Diff.Between Tests
N	27	27		
$\bar{D}$	-1.88	+0.07	t = .94	N.S.
s^2	21.32	69.97	F = 3.28	.02

This table is to be read as follows: For the 27 children in this sample, T-scores on the Raven's decreased 1.88 (negative gain) over the period Grade III to VII, while on the CTMM Language, T-scores increased an average of .07. A t-test of these gains failed to show a significant difference between the two tests. The variance of these gains (or changes) was significantly larger for the CTMM than for the Raven's.

characteristics. These characteristics are examined in relation to Hypothesis IV which suggests that certain tests will show less relationship to language ability in that they require no reading and no language of an intra-cultural type. Since a perfunctory examination of the test content reflected the factor constitution of the tests it was decided to omit any fuller discussion of the content of the tests until a later section on the evaluation of the tests. Another variable not quite as obvious as the content of the items is the amount of verbal instruction required for the administration of the tests. This aspect will also be discussed in the evaluation section.

Relationship with School Achievement.

If tests of intelligence are to be considered as valid estimates of intellectual potential, they must have some relationship to school achievement. Although some intelligence tests are validated exclusively against this criterion, for bilingual children in an adaptive treatment situation "cold-blooded" predictors are neither very valid nor are they expected to provide the best estimate of school achievement. The testing of this hypothesis is beyond the scope of this study but since the school situation is somewhat adaptive, tests of intellectual potential should have some relationship to school achievement as suggested in Hypothesis V.

The extent of this relationship was taken from Elley (1961) who attacked the problem using both average school marks and standardized tests.

The first part of this examination involved the correlating of intelligence test data with the average school mark based on teacher made tests in reading, literature, language, social studies, science, mathematics, art, music, health and personal development, and industrial arts or home economics. To minimize the effects of individual teachers' scaling procedures, correlations were calculated for each class separately and the median correlation taken to represent the relationship between each measure of intelligence and scholastic achievement as defined by the teachers. These median correlations are presented in Table X, Column A.

An examination of this column shows that the relationship of school achievement with Verbal tests, the CTMM and the Laycock is higher than that with Non-Verbal tests. However, the correlations of achievement with the Raven Matrices and the Holzinger-Crowder Series are relatively high. Considering that Verbal tests are often validated against the criterion of achievement it is to be expected that they correlate highly. On the other hand the Non-Verbal tests have reduced the overlap with achievement, hence, the relationship is striking.

TABLE X

AVERAGE CORRELATION COEFFICIENTS BETWEEN TWO MEASURES OF SCHOOL ACHIEVEMENT AND SELECTED INTELLIGENCE TESTS[1]

N = 271 Test	A Median Corr. with Average School Marks	B Mean Corr. with California Achievement
Raven Matrices	.56	.41
Cattell Test of g	.34	.35
Lorge-Thorn. Fig. Class	.31	.31
Lorge-Thorn. Series	.45	.41
Lorge-Thorn. Fig. Anal.	.42	.39
Lorge-Thorn. Total	.43	.47
Holz.-Crowder Series	.51	.49
Holz.-Crowder Fig. Changes	.40	.39
CTMM Non-Language	.38	.38
CTMM TOTAL	.66	.65
Laycock Total	.59	.64
CTMM Language	.67	.66

[1]Adapted from Elley, 1961, p. 161, 163.

The second approach to this problem was to correlate the results from the California Achievement tests given in Grade VI with the results from various intelligence tests. The mean correlation of each test of intelligence with total Reading, Arithmetic, and English Achievement scores are presented in Table X , Column B.

Again the CTMM and the Laycock show a much higher relationship with Achievement than do the Non-Verbal tests. Among the Non-Verbal tests the Holzinger-Crowder Series Test, the Lorge-Thorndike Total and Series subtest, and the Raven Matrices all show mean correlations above .4. The Cattell Test of g and the Lorge-Thorndike Figure Analogies showed slightly lower correlations than the Raven's.

The correlations of Non-Verbal tests with school achievement may have increased if the bilingual group was in a truly adaptive situation.

Elley (1961) concluded by saying:

> ...the predictive power of the culture-reduced tests is lower than that of the verbal tests, but the difference is not great enough to discourage their use with children who are handicapped verbally, particularly if measures to reduce the handicap are taken. (p. 160)

Summary. In this chapter the experimental findings pertaining to Hypotheses III, IV, and V were presented with these conclusions.

1. The Raven Matrices, a Non-Verbal test highly loaded with g provided a more stable estimate over a five year period, Grade III to Grade VII of the general intellectual potential of bilingual children than does the CTMM Language, a conventional Verbal test.

2. Tests of intelligence were examined with regard to the language requirements in both content and administration. The language requirements reflected the verbal factor loadings for each test. A more detailed analysis will be presented later.

3. The predictive power of Non-Verbal tests as indicated by the correlation of the tests with school achievement, is lower than that of the Verbal tests but the difference is not great enough to discourage their use with children from a foreign language background especially in an adaptive treatment situation. Of the Non-Verbal tests the most promising were the Raven Matrices, Holzinger-Crowder Series and the Lorge-Thorndike Total.

CHAPTER X

EVALUATION OF THE INTELLIGENCE TESTS IN THE BATTERY

The results from the testing of Hypotheses I to V will now be drawn together with a view to selecting those tests which possess a high degree construct validity as measures of the intellectual potential of children from a foreign language background for adaptive treatment situations. Table XI lists the tests examined in this study and shows the extent to which each meets the criteria specified in Postulate I.

These criteria are not all of equal importance; differential weightings were subjectively assigned to them on the basis of their relevance to the problem of validity. Since the g loading indicates the extent to which a test is measuring general intellectual ability, it was considered the most important criterion and was given the highest weighting. The second most important aspect related to this study was that the test score should be relatively independent of language background.

The other variables are related to these two major ones but being of lesser importance they received smaller weightings. Absence of group verbal and numerical factors are of some importance because they represent acquired skills and

TABLE XI

SUMMARY OF FINDINGS FOR SIXTEEN INTELLIGENCE TESTS MEASURED AGAINST SIX CRITERIA

Test	g Loadings	v:ed Loadings	n Loadings	r with L-B	Median r with School Mks.	Mean r with Cal. Ach.
Raven Matrices	.78	---	---	-.02	.56	.41
Cattell g	.79	---	---	+.04	.34	.35
L-Thorn. Fig. Class	.58	---	---	+.05	.31	.31
L-Thorn. Series	.55	---	.40	-.10	.45	.41
L-Thorn. Fig. Anal.	.74	---	---	+.15	.42	.39
L-Thorn. Total	.75	---	.33	+.07	.43	.47
Holz.-Crow. Spatial	.40	---	---	+.13	---	---
Holz.-Crow. Series	.46	.21	.45	-.14	.51	.49
Holz.-Crow. Fig. Cl.	.52	---	.21	-.20	.40	.39
CTMM Spatial	.61	---	---	-.03	---	---
CTMM Logical	.66	---	---	+.34	---	---
CTMM Numerical	.64	.20	.33	-.19	---	---
CTMM Verbal	.46	.66	---	+.26	---	---
CTMM Lang. Total	.58	.62	---	+.26	.67	.66
CTMM Non-Lang. Total	.62	---	.35	-.07	.38	.38
CTMM Total	---	---	---	+.16	.66	.65
Laycock	.63	.57	---	+.33	.59	.64

are not a direct function of general intellectual potential. Stability over a four year period is of lesser importance in evaluation because of the small discrepancy between tests and also because not all tests could be examined against this criterion. The test content should depend very little on acquired information and be equally unfamiliar to all testees. Administration of the test should also minimize the dependency on English instructions for children from a foreign language background. Medium correlations (approximately .5) with school achievement are necessary since achievement should be related to general intellectual ability.

In order to make the rating process more explicit specific numerical weightings were assigned to each of these criteria as shown in the top row of Table XII.

Each test was then given a rating out of a maximum of five points on each of these criterion variables. The loading of each test was multiplied times the weighting of the criterion and these were then summed over all criteria to yield a final value for the test. Loadings were substituted in the following equation:

$$6a - 3b - 2c - 5d + 2e - 1f - 2g + 3h = \text{value of test}$$

These loadings and the obtained values of each of the tests as measures of the intellectual potential of bilingual

TABLE XII

EVALUATION OF SIXTEEN TESTS MEASURED AGAINST EIGHT CRITERIA

Test	g Loading	v:ed Loading	n Loading	r with Lang.Bk.	Stab-ility	Dependency on English Admin.	Dependency on English Cont.	Correl. with Achiev.	Total Final Evaluation
Weighting	6	-3	-2	-5	2	-1	-1	3	
Raven Matrices	4	0	0	0	3	0	0	3	39*
Cattell g	4	0	0	0	3	2	0	3	37*
L-Thorn. Fig. Class.	3	0	0	0	3	2	0	2	28
L-Thorn. Series	3	0	3	1	3	2	2	3	18
L-Thorn. Fig. Anal.	4	0	0	1	3	2	0	3	32*
L-Thorn. Total	4	0	3	0	3	2	1	4	33*
Holz.-Crow. Spatial	2	0	0	1	3	2	0	2	17
Holz.-Crow. Series	2	2	3	1	3	2	2	4	9
Holz.-Crow. Fig. Changes	2	0	2	2	3	2	0	3	11
CTMM Spatial	3	0	0	0	3	2	0	2	28
CTMM Logical	3	0	0	3	3	2	3	2	10
CTMM Numerical	3	2	3	2	3	2	2	2	4
CTMM Verbal	2	4	0	2	3	2	3	4	3
CTMM Language	3	4	0	2	2	2	3	4	7
CTMM Non-Language	3	0	3	0	3	2	1	3	24
CTMM Total	3	3	2	2	2	2	3	4	6
Laycock	3	4	0	3	3	2	4	4	3

*Tests with highest value as measures of general intellectual ability for bilingual children.

children are given in Table XII.

An examination of the total values assigned to each of the tests shows that they form an array from a maximum value of thirty-nine assigned to the Raven Matrices to a minimum of three assigned to the Laycock. The highest scores represent greater validity in the intelligence testing of bilingual children.

The four tests which emerge as the most promising instruments are the Raven Matrices, the Cattell Test of g, and the Lorge-Thorndike Figure Analogies and Total Test. Other instruments which should not be overlooked are the Lorge-Thorndike Figure Classification and the CTMM Non-Language.

The Raven Matrices is one of the best single instruments evaluated in this study. It has a high g loading, no loadings over .20 on the v:ed or n factors, has little relationship to language background, has shown itself to be relatively stable over a four year period and shows medium correlations with achievement. It is also outstanding in that it has little cultural or acquired content and it requires almost no verbal instructions in administration.

The Cattell Test of g has very similar characteristics to that of the Raven's with the exception that the administration of the test requires a good understanding of English. Elley (1961) suggests that with slight modification in the inst-

ructions this instrument could be even more effective than it already is.

The Lorge-Thorndike Figure Analogies subtest must be ranked as one of the most effective instruments in the battery. It met nearly all the criteria as well as the Raven and Cattell in spite of the fact that it takes only nine minutes to administer. The Lorge-Thorndike Total has little virtue beyond the shorter figure Analogies subtest, moreover, it has a loading on the number factor which must be construed as a limitation. In spite of this, it must be ranked as one of the best instruments available for the testing of verbally handicapped children.

The remaining tests and subtests fail to meet one or more of the criteria and hence obtain lower total scores on the ratings. Verbal tests, as expected, are near the bottom of this rating scale, in other words, these instruments are not as suitable for the intellectual assessment of children from a foreign language background.

Summary. An evaluation of the sixteen intelligence tests in the battery against eight weighted criteria led to the conclusion that the Raven Progressive Matrices, the Cattell Test of g, and the Lorge-Thorndike Figure Analogies were the most promising instruments for assessing the intellectual potential of children with a foreign language background. Two other tests, the Lorge-Thorndike Figure Classification and the CTMM Non-

Language, while markedly inferior for this purpose to the first three tests, were found to be of intermediate effectiveness. Verbal tests were found to be less valid than Non-Verbal tests for estimating the intellectual ability of bilingual children.

CHAPTER XI

DEGREES OF FOREIGN LANGUAGE AS RELATED TO INTELLIGENCE

The "B" section of the hypotheses were included for a purpose quite different from that of part "A". The first section dealt with the selection of intelligence tests with construct validity as measures of the intellectual ability of bilingual children. The second part is an attempt to see the effect of varying degrees of foreign language background on intelligence test scores. However, some of the calculations in Part A are relevant to this discussion.

As the review of the literature indicated one would hypothesize that Verbal intelligence scores would be a decreasing monotonic function of the degree of bilingualism while the degree of bilingualism would have little or no effect on Non-Verbal tests of intelligence. Since Edmonton is not a truly bilingual community and since there are a variety of languages and not just two, one does not expect to find precisely the same relationship; that a similar relationship obtains is suggested by Hypotheses VI and VII.

Degree of Foreign Language Background

In order to ascribe the differences between groups to language background it was necessary to exclude those variables

that may confuse or bias the results, Hence, the groups were matched on socio-economic status while age considerations were brought into the interpretations of the results.

Evidence of this relationship is presented in three ways: comparison of mean T-scores, comparison of mean IQ's and by graphical representation.

Table XIII shows the mean T-scores on various intelligence tests for the three groups each with a different degree of foreign language background. Figure 2 is a graphical representation of these mean T-scores.

An inspection of this table and its associated graph shows that the general trend found in other studies is maintained here. Bilingual scores are significantly lower than unilinguals on Verbal tests while on Non-Verbal tests no differences were apparent.

This comparison is limited by the fact that the groups do not have equal mean ages, the unilingual group being seven months younger than the bilingual with the middle language group in between. Since intelligence test scores increase with age, the superiority of the unilingual group would be even more pronounced on Verbal tests if the mean age of this group was increased by seven months. Why matching of groups on age was not done was discussed in Chapter VI.

A crude attempt to take this age differential into account was made by converting the mean raw score on each test to a mean IQ for each group by using the mean chronological age and

TABLE XIII

MEAN T-SCORES ON SELECTED INTELLIGENCE TESTS FOR THREE GROUPS FORMED ON THE BASIS OF DIFFERENT HOME LANGUAGE BACKGROUND

Test	Group I Uniling.	Group II	Group III Bilingual	Sign. of Difference I - III
N	40	33	22	
Raven Matrices	48.68	49.40	48.45	N.S.
Cattell Test of g	47.79	47.42	47.05	N.S.
Lorge-Thorn. Total	47.36	46.60	46.05	N.S.
CTMM Non-Language	47.13	49.29	48.52	N.S.
CTMM Language	48.11	47.39	42.76	.05
Laycock	49.03	45.53	42.90	.01

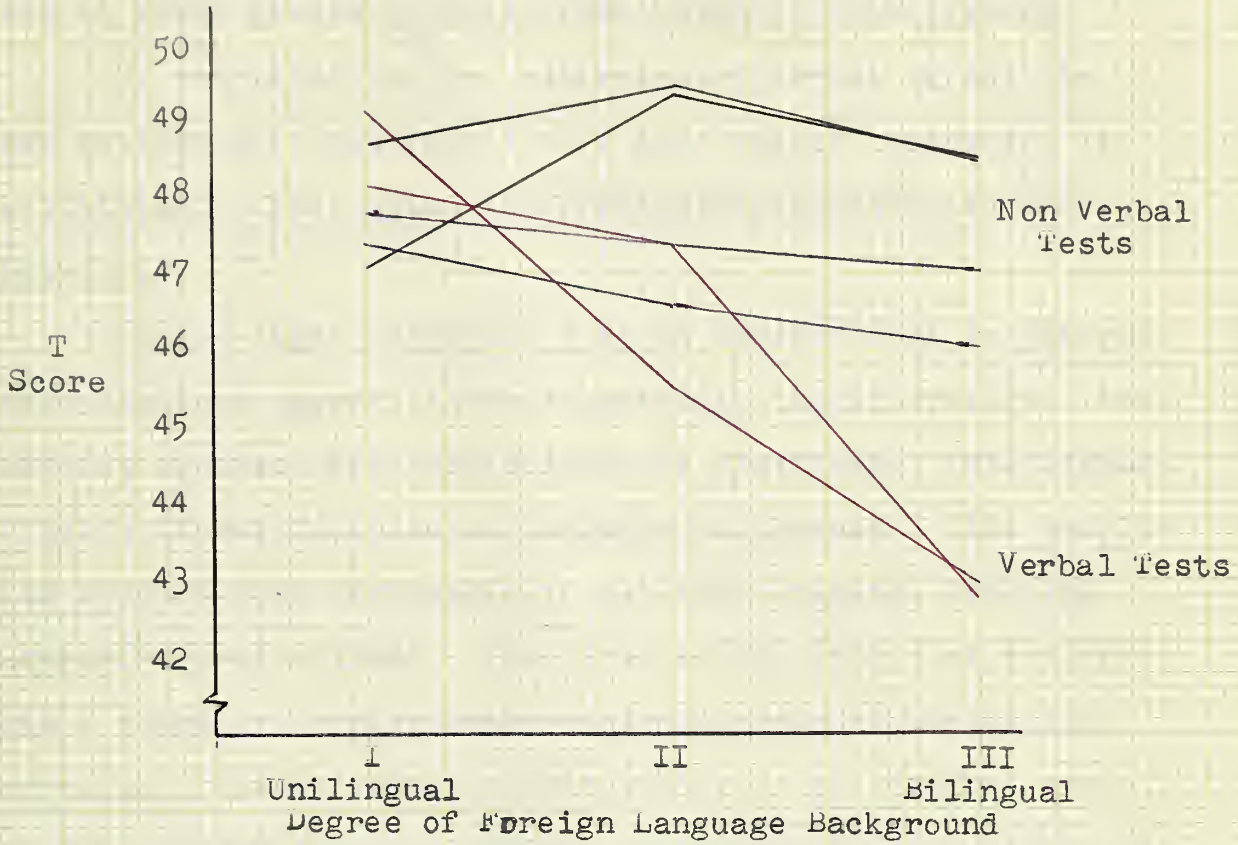

FIGURE 2

CURVE SHOWING THE RELATIONSHIP BETWEEN DEGREE OF FOREIGN LANGUAGE BACKGROUND AND MEAN T-SCORE ON SELECTED INTELLIGENCE TESTS

the publisher's norms. Since these IQ's are based on publishers norms they are not strictly comparable, only the trends are dependable. Table XIV presents the mean IQ's for each of these groups on these same tests of intelligence.

By combining the two estimates of Verbal IQ and the four estimates of Non-Verbal IQ a more stable comparison is facilitated. These means are graphically represented in Figure 3.

In the light of Figure 3 it is obvious that in general these findings support those reported in the literature. One fact not apparent from the diagram is that Verbal intelligence is not a linear function of language background. This conclusion becomes more noticeable if all four original language categories are included. The first of the groups of children with a slightly foreign background was excluded because the scores were almost identical with those of the control group. In other words, foreign language background of this slight degree actually had no effect on intelligence test scores. Figure 4 shows the nature of the relation between these two variables.

From an examination of Figure 4 it may be tentatively concluded that intelligence as measured by Verbal tests is a positively decelerated function of the degree of foreign language background and not a linear one as suggested earlier. If this is the case it is easy to understand why children from a

TABLE XIV

MEAN IQ'S ON SELECTED INTELLIGENCE TESTS FOR THREE GROUPS FORMED ON THE BASIS OF DIFFERENT HOME LANGUAGE BACKGROUND

Test	Group I Uniling.	Group II	Group III Bilingual
N	40	33	22
Raven Matrices	105	104	102
Cattell Test of g	105	102	100
Lorge.-Thorn. Total	109	106	104
CTMM Non-Language	110	109	106
Average Non-Verbal IQ	107	105	103
CTMM Language	112	109	102
Laycock	98	91	86
Average Verbal IQ	105	100	94

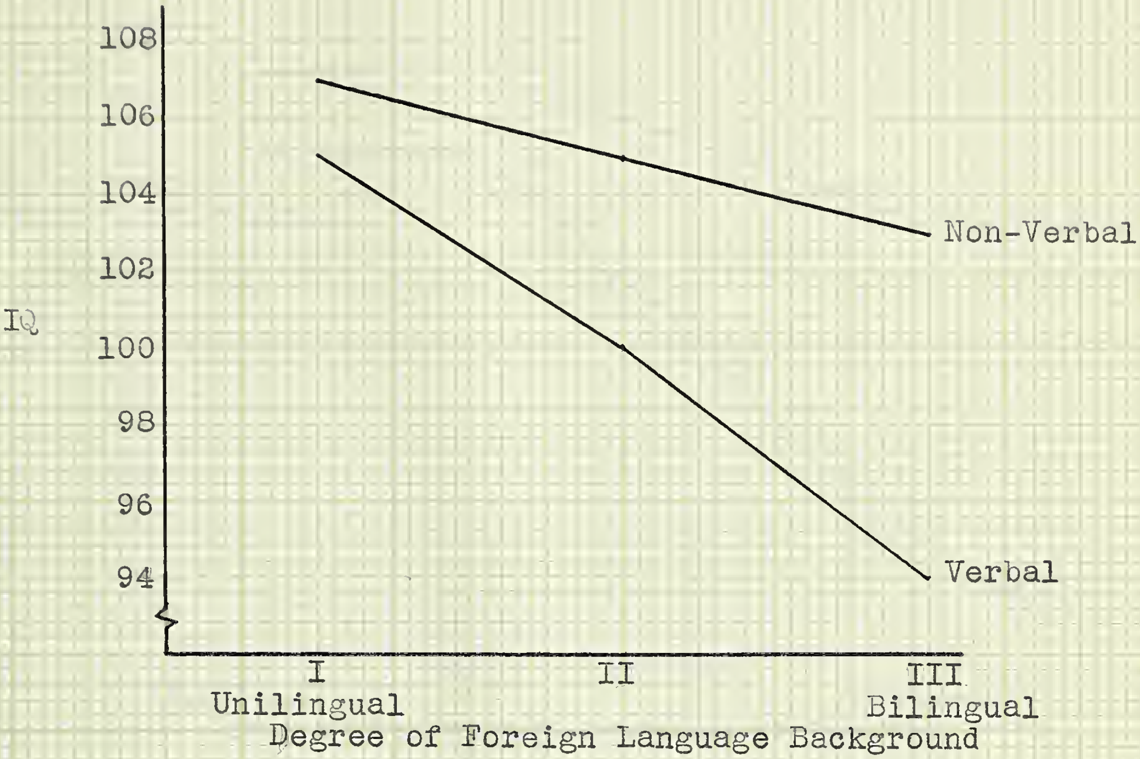

FIGURE 3

CURVE SHOWING THE RELATIONSHIP BETWEEN DEGREE OF FOREIGN LANGUAGE BACKGROUND AND MEAN IQ FOR VERBAL AND NON-VERBAL TESTS

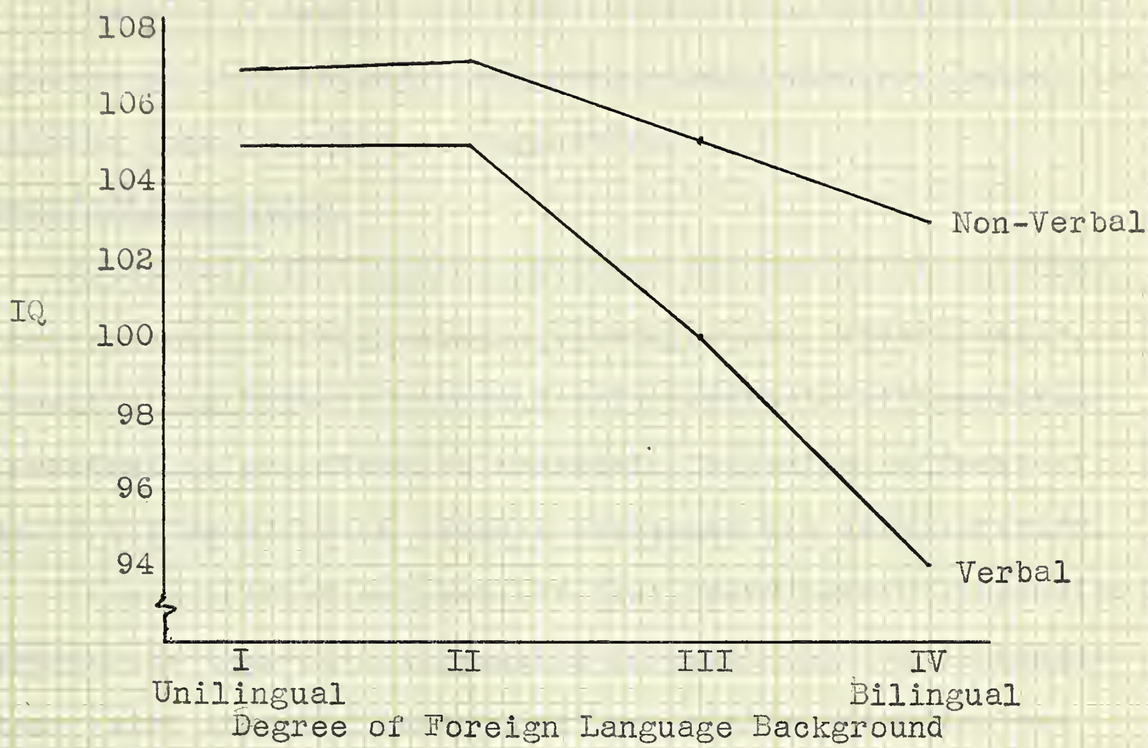

FIGURE 4

CURVE SHOWING THE RELATIONSHIP BETWEEN FOUR CATEGORIES OF INCREASING DEGREE OF FOREIGN LANGUAGE BACKGROUND AND MEAN IQ ON VERBAL AND NON-VERBAL TESTS

slightly or moderately bilingual background score only slightly if any, lower than the unilingual children on most tests of intelligence.

There is a danger that this curve is an artifact due to the nature of the Language Background questionnaire, hence, the conclusion must be taken as suggestive.

Type of Bilingualsim

Hypothesis VIII sought to tease out one of the nuances of the variable, foreign language background. Specifically, a differentiation between compound and coordinate bilingualism was examined in an effort to ascertain the actual effects of types of bilingualism on thought processes via verbalization.

It was pointed out earlier that coordinate bilingualism refers to the type of bilingualism in which the two language systems exist somewhat independently; in this situation one expects a minimal amount of interference in language and thought process.

If a child learned the two languages at a different age and in a different setting he was presumed to be a coördinate bilingual. On the other hand, if the child was born in Canada and learned both languages concurrently, he was classified as a compound bilingual.

A preliminary examination revealed that for the data of this study, the dichotomy was untestable. It was found that

the groups set up on the basis of compound versus coordinate bilingualism were almost identical with those set up on the basis of immigrant versus second-generation Canadians. That is, all immigrants were classified as coordinates while nearly all second-generation Canadians were classified as compound. It was therefore impossible to attribute differences between these groups to either set of variables. Nevertheless, an analysis of covariance was carried out to see if this dichotomy based on the number of years in Canada and the age at which English was learned had any effect on Verbal test score beyond that attributable to degree of foreign language background. (Garrett, 1958, p. 295) The Laycock Mental Ability test was employed because it was the most likely to show significant differences. The results of the analysis of variance of language background and Laycock score taken separately are shown in Table XV .

This table shows a significant difference between the groups on language background and the difference between groups on the Laycock approaches significance. The effect of removing the influence of language background by an analysis of covariance is shown in Table XVI.

This table shows that once the effect of Language Background score is removed, no differences in intelligence remained between these groups.

TABLE XV

ANALYSIS OF VARIANCE OF LANGUAGE-BACKGROUND (X)
AND LAYCOCK INTELLIGENCE SCORE (Y)
TAKEN SEPARATELY

Source of Variation	d.f.	SS_x	SS_y	V_x	V_y
Between Means	1	2	525	2.0	525
Within Groups	49	10	10341	.2	211
Total	50	12	10866		

$F_x = \frac{2}{.2} = 10$ Significant at .01 level

$F_y = \frac{525}{211} = 2.49$ Not Significant

From table of F for d.f. 1/49

F at .05 level = 4.03

F at .01 level = 7.17

TABLE XVI

ANALYSIS OF COVARIANCE SHOWING THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN GROUPS WITH THE DEGREE OF FOREIGN LANGUAGE BACKGROUND REMOVED

Source of Variation	d.f.	$SS_{y \cdot x}$	$V_{y \cdot x}$
Between Means	1	250	250
Within Groups	49	10340	207
Total	50	10590	

$F_{y \cdot x} = \frac{250}{207} = 1.21$ Not Significant

From table of F for d.f. 1/49

F at .05 level = 4.03

F at .01 level = 7.17

Therefore, for parsimonious explanation of the data the variables of compound versus coordinate or immigrant versus second-generation Canadian (as implied through data on the age at which English was learned) were omitted and the results accounted for simply in terms of degree of foreign language background.

Summary. On the basis of the scores of Edmonton Grade VII children on various intelligence tests it was concluded that while Non-Verbal intelligence is relatively independent of language background, Verbal intelligence is a function of the degree of foreign language background. Rather than a linear relationship, however, the results suggest a positively decelerated one. That is, a slight degree of foreign language background has a negligible effect but a higher degree of foreign background has an effect on Verbal test scores.

It was also concluded that for the data of this study the variables compound versus coordinate and immigrant versus non-immigrant were unnecessary. A more parsimonious explana - tion of the effect of a foreign language home background can be given in terms of the score on the Language Background Questionnaire.

CHAPTER XII

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

The purpose of this study was to ascertain the effects of a foreign language background on selected intelligence tests with a view to selecting tests which provide a valid estimate of the general intellectual ability of children with a foreign language background. As such, this study is parallel to that of Elley (1961) in which he selected tests of general intellectual ability with a minimum of socio-economic bias.

The problem was attacked by analyzing the results of nine "culture-reduced" tests and two conventional Verbal intelligence tests administered to 432 Edmonton Grade VII children. On the basis of the Language Background Questionnaire subjects were classified into three groups, representing increasing degrees of foreign language background, matched on socio-economic status with N's of forty, thirty-three and twenty-two respectively.

For a sample of 271 of these children representative of the Grade VII pupils of Edmonton Public Schools, Elley (1961) used the results from twenty-one tests in a Principle Components factor analysis. Since intelligence was regarded theoretically as a hierarchy of abilities, a high loading on the broad general intellectual ability factor g was taken as the major criterion in the selection of valid in-

struments. Moreover, since the literature points out that one of the major determiners of the intelligence scores of bilingual children is the nature of the test, it is essential to select tests which measure the most basic aspect of intelligent behavior, namely, general intellectual ability.

The other major criterion for construct validity was the independence of the intelligence test and foreign language background. This independence was examined by comparisons of tests across language groups, by comparisons of Non-Verbal with conventional tests and by correlations of tests with language background.

The findings from these procedures were then combined to evaluate each test as a measure of the intellectual potential of bilingual children.

The final aspect of this study deals with the nature of the relationship between different degrees of bilingualism and intelligence test scores.

Experimental Findings

1. Factor analysis of the twenty-one variables produced one large general intellectual factor g which accounted for over fifty percent of the common variance, three group factors which were labelled verbal-educational (v:ed), numerical (n), speed (sp) and a socio-economic factor. The tests most adequately meeting the criterion of high loading on g

with negligible loadings on group factors were the Raven Matrices, Cattell Test of g and the Lorge-Thorndike Total.

2. Two conventional Verbal tests, the California Test of Mental Maturity and the Laycock showed significant differences between the unilingual and bilingual groups whereas the Non-Verbal tests showed no significant differences. Moreover, by correlating language background with each intelligence test similar results were obtained. On the basis of these two approaches the tests emerging with the least relationship to language background were the Raven Matrices, Cattell Test of g, Lorge-Thorndike Figure Classification and the CTMM Spatial.

3. When a group with a moderate degree of foreign language background was compared with the unilingual group no significant differences were found. The mean scores on Verbal intelligence tests tended to be slightly lower for this group than for the unilingual group.

4. In an analysis of the changes in T-scores over a four year period on each of two intelligence tests no conclusions could be reached on the direction of the changes. However, on the basis of the variance of the gains it was concluded that for bilingual children, scores on the Raven's, a Non-Verbal test, highly loaded with g, remain more stable over a four year period than the CTMM Language, a conventional Verbal test.

5. Verbal intelligence tests appear to make heavy de-

mands on culturally acquired information including language; Non-Verbal tests generally minimize this demand. All the tests except the Raven Matrices make some demand on verbal facility in understanding administrative instructions.

6. Verbal intelligence tests showed higher correlations with school achievement than did Non-Verbal tests. In view of the somewhat academic content of conventional intelligence tests and the fact that they are often validated against academic criteria, this higher relationship is to be expected. However, Non-Verbal tests, particularly the Holzinger-Crowder Series, the Lorge-Thorndike Total and the Raven's correlated sufficiently well to warrent their use in academic prediction for bilingual children. It is likely that the predictive power would increase in an adaptive treatment situation.

7. Verbal intelligence scores were found to be inversely related to language background scores while Non-Verbal scores were relatively independent of language background.

Conclusions

1. On the basis of eight criteria used in this investigation it was concluded that the Raven Standard Progressive Matrices was the most adequate instrument for the intellectual assessment of bilingual children. This test rated satisfactorily on the following criteria:

a. high g loading,

b. absence of group factors,

c. little relationship to language background,

d. relative stability for bilingual children over the period Grade III to VII,

e. lack of dependence on English in both content and administration, and,

f. satisfactory correlations with achievement.

In addition to its validity this test is unspeeded, intrinsically interesting, easily administered, and reliable. This conclusion is similar to that reached by Morgan (1957, p. 15) in a study of intelligence testing in a bilingual area in Wales: "Of the three tests employed, Raven's Progressive Matrices (1938), the only untimed test, proved to be the most independent of Welsh linguistic background and the most reliable measure of non-verbal intelligence [although]..... none of the tests employed can be regarded as wholly satisfactory." Moreover, Elley (1961) has shown the Raven's to be the most useful instrument for the intellectual assessment of children from low socio-economic areas, a problem similar to that of the present study.

The Cattell Test of g met all the criteria as effectively as the Raven's except for the difficulty in interpreting the highly verbal directions of the Cattell. For a child with a fair understanding of English this limitation is minimized.

The Lorge-Thorndike Figure Analogies subtest was almost

as effective as the other two tests, its limitation being its slight relationship with a foreign language background.

Any one of these three tests could be used more effectively for assessing the intellectual potential of children from a wide range of foreign language backgrounds than most conventional intelligence tests.

2. The reported relationship between degree of bilingualism and scores on Verbal and Non-Verbal test was confirmed in this study. Whereas Non-Verbal intelligence tests are relatively independent of the degree of language background, verbal test scores are inversely related to language background scores. The former supports the contention of Jones (1959, p. 46) that "bilingualism is not necessarily a source of intellectual disadvantage". Evidence suggests that the latter curve may be negatively accelerated in that a mild degree of bilingualism has little or no effect on Verbal tests.

Implications

1. Theoretical. The hypotheses for this study were extentions of a series of postulates which constitute the theoretical framework. Confirmation of the hypotheses adds credence to the postulates; rejection of the hypotheses necessitates some changes.

In general, the theory advanced at the outset of this study has been substantiated by the experimental results.

a. Bilingual children from a predominently non-English home tend to be limited in verbal ability in English;

this limitation is manifested by lower scores on Verbal tests of intelligence. However, intellectual functioning is apparently not impaired in that Non-Verbal tests requiring general intellectual ability fail to discriminate across groups.

b. Tests intended to measure the intellectual ability of bilingual children should minimize the language content while maximizing the abstract reasoning ability as indicated by the g loading.

c. Tests are available which provide a useful estimate of the general intellectual ability of children from a wide variety of backgrounds.

Two aspects of the theory need to be changed somewhat.

a. Postulate 5a suggests a linear relationship between language background and Verbal intelligence. It was assumed therefore, that a slight foreign language background would have a slight inhibitory effect on Verbal intelligence. Although this may be the case for truly bilingual communities, it appears that for this Edmonton sample, a slight or even moderate foreign language background has no effect on Verbal intelligence. It is only when the degree of bilingualism becomes quite high, perhaps as high as using a language other than English one-half or more of the time, that it has sufficient effect on verbal ability to actually interfere with the assess-

ment of intelligence by conventional tests.

b. The dichotomy between compound and coordinate bilingualism may be either irrelevant or premature because of the absence of a truly bilingual community and because of the crude measures of intellectual functioning.

In general, this study supports the contention that g is a useful construct and that it can be assessed relatively independent of verbal or other cultural factors.

2. Practical. It would appear profitable to employ one of the three Non- Verbal tests of intelligence validated by this study for the classification and grouping of bilingual children. Moreover, selection of children for intensive adaptive treatment could be based to some extent on results from these tests. Finally, experiments requiring an estimate of intellectual potential should make use of these Non-Verbal tests.

The findings of this study have further significance in that they have undermined a longstanding assumption regarding the mental inadequacy of children from non-English homes. Children from homes in which a language other than English is spoken only a small part of the time are as effective as unilinguals on both Verbal and Non-Verbal tests of intelligence.

3. Implications for Further Research. This writer is of the opinion that sufficient evidence regarding the potential

usefulness of the statistical factor g and the possibility of its adequate measurement, has been accumulated; little is to be gained from persuing this approach for the present. The virtue of an estimate of intellectual potential although fairly well established remains almost completely hypothetical. Evidence now needs to be obtained regarding the empirical and pragmatic usefulness of measures of the construct g. Studies similar to that now being carried out by the New York Board of Education (1961) need to be carried out in Alberta in order to determine the feasibility of adaptive treatment and its relationship to general intellectual potential. _If_ it can be demonstrated that adaptive treatment is more effective for children of high intellectual potential, and _if_ it can be demonstrated that measures of intellectual potential have superior predictive power prior to the adaptive treatment, then further refinement of these measures is in order. Until such a time further refinement and analysis is premature.

However, certain less crucial elements could be usefully explored.

a. A factor analysis of these tests for a group composed largely of children with a foreign language background may result in modified loadings of each of the tests on the factors. For example, the CTMM Logical may be expected to exhibit a larger verbal loading.

b. Improvements in the Language Background Questionnaire may lead to a more effective examination of the relationship between language background and intelligence. The suggestion of a negatively accelerated curve needs further examination.

c. An examination of the extent to which different languages fit the curves representing the general class of foreign language background may lead to a clearer more precise description of the relation of this variable to intelligence.

Finally a word of warning to those conducting similar studies. An avoidable problem was introduced into this study because the sampling had been done by grade rather than by age. Sampling by grade produced a negative correlation between intelligence test score and age, a phenomenon which made experimental outcomes difficult to interpret.

BIBLIOGRAPHY

Altus, Grace, "WISC Patterns of a Selective Sample of Bilingual School Children." J. Genetic Psych., 1953, 83, 241-248.

Anastasi, A. and Cordova, F. A., "Some Effects of Bilingualism Upon the Intelligence Test Performance of Puerto Rican Children in New York City." J. Educ. Psych., 1953, 44, 1-19.

Arsenian, Seth, "Bilingualism in the Post-War World." Psych. Bull. 1945, 42, 65 - 86.

Barke, E. M., and Williams, D. E. P., "A Further Study of the Comparative Intelligence of Children in Certain Bilingual and Monoglot Schools in South Wales." Br. J. Educ. Psych., 1938, 8, 63 - 77.

Bilingualism - A Bibliograph with Special Reference to Wales, Pamphlet No. 7. Aberystwyth, University College of Wales, 1960.

Blishen, B. R., "The Construction and Use of an Occupational Class Scale." Can. J. Econ. and Pol. Science. 1958, 24, 519 - 531.

California Achievement Tests: Primary. Los Angeles:

California Test Bureau, 1951.

California Achievement Tests: Junior High Level. Los Angeles: California Test Bureau, 1957.

California Short-Form Test of Mental Maturity: Primary. Los Angeles: California Test Bureau, 1953.

California Short-Form Test of Mental Maturity: Elementary. Los Angeles: California Test Bureau, 1957.

Carlson, H. B. and Henderson, N. "The Intelligence of American Children of Mexican Parentage." J. Abnorm. and Soc. Psych., 1950, 45, 544-551.

Cattell, R. B., Handbook for Culture - Free Intelligence Test. Scale 2. Champaigne, Illinois: I.P.A.T., 1959.

Chauncey, H., Educational Testing Service. Annual Report of the President, 1958-59, Princeton.

Clark, E. L., "The Motivation of Jewish Students." J. Soc. Psych., 1949, 29, 113-117.

Colvin, S. S. and Allen, R. D., "Mental Tests and Linguistic Ability." J. Educ. Psych., 1923, 14, 1-20.

Compulsory Welsh. Times Educ. Supp. 2191:694 May 17, 1957.

Cooper, J. G., "Predicting School Achievement for Bilingual Pupils." J. of Educ. Psych., 1958, 49, 31-36.

Coull, W. H., A Normative Survey of Reading Achievement of Alberta Children in Relation to Intelligence, Sex, Bilingualism, and Grade Placement. Unpublished M. E. D. Thesis, Univ. of Alberta, 1956.

Cronbach, L. J. and Gleser, G.C., Psychological Tests and Personnel Decisions. Urbana, Univ. of Ill. Press, 1957.

Cronbach, L.J., Essentials of Psychological Testing. New York: Harper and Bros., 1960.

Darcy, N.T., "The Effect of Bilingualism upon the Measurement of the Intelligence of Children of Pre-School Age." J. Educ. Psych., 1946, 37, 21-44.

Darcy, N.T., "The Performance of Bilingual Puerto Rican Children on Verbal and Non-Language Tests of Intelligence." J. Educ. Res., 1952, 45, 499-506.

Darcy, N. T., "The Effect of Bilingualism upon the Measurement of Intelligence - a Review of the Literature." J. Genetic Psych., 1953, 82, 21-57.

Elley, W. B., A Comparative Analysis of the Socio-Economic Bias in Selected Intelligence Tests. Unpublished Ph. D. Thesis, University of Alberta, 1961.

Ellis, E. N., The Effectiveness of Culture-Free Tests in Measuring the Intellectual Characteristics of German Immigrants to Canada. Abstract of Unpublished Ed. D. Thesis, Oregon State University, 1956.

Ferguson, G.A., Statistical Analysis in Psychology and Education. Toronto: McGraw - Hill, 1959.

Gardner, R.C. and Lambert, W.E., "Motivational Variables in Second Language Acquisition." Can. J. Psych., 1959, 13, 266-272.

Garrett, H. E., Statistics in Psychology and Education. New York: Longmans, Green and Co., 1958,

Guilford, J. P., "The Structure of Intellect." Psych. Bull., 1956, 53, 267-293.

Harman, H. H., Modern Factor Analysis. Toronto: University of Toronto Press, 1960.

Hartmann, Sr. Mary Andrew, "Variables Considered in the Measurement of Bilingualism - a Review of the Literature." Revue de L'Universite d'Ottawa, 1961.

Havighurst, R. J. and Hilkevitch, R. R., "The Intelligence of Indian Children as Measured by a Performance Scale." J. Abnorm. and Soc. Psych., 1944, 39, 419-433.

Hebb, D. O., A Textbook of Psychology. Philadelphia: W. B. Saunders Co., 1958.

Hoffmann, M., The Measurement of Bilingual Background. New York: Teachers' College, Columbia University, 1935.

Holzinger, K. J. and Crowder, N. A., Holzinger-Crowder Uni-Factor Tests. New York: World Book Co., 1955.

Hunt, J. McV., Intelligence and Experience. New York: Ronald Press, 1961.

James, C. E. B., "Bilingualism in Wales: an Aspect of Semantic Organization." Educ. Res., 1960, 2, 123-126.

Jamieson, E. and Sandiford, P., "The Mental Capacity of Southern Ontario Indians." J. Educ. Psych., 1928, 19, 313-328.

Johnson, G. B., "The Relationship Existing Between Bilingualism and Racial Attitude." J. Educ. Psych., 1951, 42, 357-65.

Jones, W. R., "The Influence of Reading Ability in English on the Intelligence Test Scores of Welsh-Speaking Children." Br. J. of Educ. Psych., 1953, 23, 114-120.

Jones, W. R., Bilingualism and Intelligence. Cardiff: Univ. of Wales Press, 1959.

Jones, W. R., "A Critical Study of Bilingualism and Non-Verbal Intelligence." Br. J. of Educ. Psych., 1960, 30, 71-77.

Keston, J. J. and Jimenez, C., "A Study of the Performance on English and Spanish Editions of the Stanford-Binet Intelligence Test by Spanish-American Children." J. Genetic Psych., 1954, 85, 263-269.

Kittell, J. E., "Bilingualism and Language - Non-Language Intelligence Scores of Third-Grade Children." J. Educ. Res., 1959, 52, 263-268.

Lambert, W. E., Havelka, J., and Crosby, C., "The Influence of Language Acquisition Contexts on Bilingualism." J. Abnorm. and Soc. Psych., 1958, 56, 239-244.

Lambert, W. E., and Fillenbaum, S., "A Pilot Study of Aphesia among Bilinguals." Can. J. Psych., 1959, 13, 28-34.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., and Fillenbaum, S., "Evaluational Reactions to Spoken Languages." J. Abnorm. and Soc. Psych., 1960, 60, 44-51.

Laycock, S. R., Manual of Directions: Laycock Mental Ability Test. Saskatoon: Univ. of Sask. Bookstore, 1935.

Levinson, B. M. "A Comparative Study of the Verbal and Performance Ability of Monolingual and Bilingual Native Born Jewish Preschool Children of Traditional Parentage." J. Genetic Psych., 1960, 97, 93-112.

Lewis, D. G., "Bilingualism and Non-Verbal Intelligence: A Further Study of Test Results." Br. J. of Educ. Psych. 1959, 29, 17-22.

Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston: Houghton Miffin Co., 1953.

Lorge, I. and Thorndike, R. L., Lorge-Thorndike Intelligence Test. Boston: Houghton Miffin Co., 1957.

MacArthur, R. S., "The Colored Progressive Matrices." Alberta J. of Educ. Res., 1960, 6, 67-75.

MacArthur, R. S., Assessing General Intellectual Ability for Adaptive Teaching Decisions. Paper presented at Annual Meeting of Can. Psych. Assoc., Montreal, 1961.

Manuel, H. T., "Bilingualism". In Encyclopedia of Educational Research, 1960, 146-150.

McCarthy, D. "Language Development in Children." In Carmichael, L., Manual of Child Psychology, New York: Wiley and Sons, 1954.

McClleland, D. C., Baldwin, A. L., Bronfenbunner, V. and Strodbeck, F. L., Talent and Society. New York: Van Nostrand Co., 1958.

Meehl, P. E. and McCorquordale, K., "On a Distinction between Hypothetical Construct and Interveening Variables." Psych. Rev. 1948, 55, 95-107.

Mitchell, A. J., "The Effect of Bilingualism in the Measurement of Intelligence." Elem. Sch. J., 1937, 38, 29-37.

Morgan, E. R., Bilingualism and Non-Verbal Intelligence: A Study of Test Results. Pamphlet No. 4. Aberystwyth: Collegiate Faculty of Education, 1957.

Murdoch, K., Maddow, D. and Berg, N. L., "A Study of the Relation between Intelligence and the Acquisition of English." In National Society for the Study of Education. Twenty-seventh Yearbook, 1928.

Osgood, C. E., "Psycholinguistics - A Survey of Theory and Research Problems". Supplement to J. Abnorm. and Soc.Psych., 1954, 49.

Pintner, R. and Keller, R., "Intelligence of Foreign Children". J. Educ. Psych., 1922, 13, 214-222.

Pintner, R., "Comparison of American and Foreign Children on Intelligence Tests". J. Educ. Psych., 1923, 14, 292-295.

Pintner, R. and Arsenian, S., "The Relation of Bilingualism to Verbal Intelligence and School Adjustment." J. Educ. Res., 1937, 31, 255-263.

Raven, J. C. Guide to Using the Progressive Matrices. London: Lewis and Co., 1954.

Reid, T. J. and Conquest, G. R. "A Survey of the Language Achievement of Alberta School Children." Alberta J. of Educ. Res., 1955, 1, 39-52.

Sanchez, G. I., "Group Differences and Spanish Speaking Children- A Critical Review." J. App. Psych., 1932, 16, 549-558.

Sanchez, G. I., "Bilingualism and Mental Measures". J. App. Psych., 1934, 18, 765-772.

Singer, H. "Bilingualism and Elementary Education." <u>Mod. Lang. J.</u>, 1956, 40, 444-458.

Soffietti, J. P., "Bilingualism and Biculturalism". <u>J. Educ. Psych.</u>, 1955, 46, 222-227.

Spearman, C. E. <u>The Nature of Intelligence and the Principles of Cognition</u>. London: MacMillan and Co., 1923

Spearman, C. E. <u>Abilities of Man.</u> London: MacMillan and Co., 1932.

Spoerl, D. T., "Bilinguality and Emotional Adjustment". <u>J. Abnorm. and Soc. Psych.</u>, 1943, 38, 37-57.

Stark, W. A., "The Effect of Bilingualism on General Intelligence: An Investigation Carried on in Certain Dublin Primary Schools." <u>Br. J. of Educ. Psych.</u>, 1940, 10, 78-79.

Tireman, L. S., "Bilingual Children". <u>Review of Educ. Res.</u>, 1941, 11, 340-352.

Terman, L. M., "The Vocabulary Test as a Measure of Intelligence". <u>J. Educ. Psych.</u>, 1918, 9, 452-466.

Vernon, P. E. <u>The Structure of Human Abilities.</u> London: Methven and Co., 1950.

Walker, H. and Lev, J. <u>Statistical Inference</u>. New York: Holt, 1953.

Woodrow, H., "The Relation between Abilities and Improvement with Practice." <u>J. Educ. Psych.</u>, 1938, 29, 215-230.

Wrightstone, J. W. and Lazar, M., <u>Demonstration Guidance Project--Progress Report.</u> New York: Board of Education, 1957.

APPENDIX A

LANGUAGE BACKGROUND

Name: ______________________ Age: - __________ Grade: __________

School: ______________________ Date of Birth: ______________________

DIRECTIONS:

In order to understand and help children who speak two or more languages, we are trying to find the way that languages are used in people's homes.

You can help us by putting an [X] in the right boxes below.

PART I

	Never	Small Part of the Time	Most of the Time	All of the Time
EXAMPLE: How often do you whisper in school?				
1. How often do any of these people speak to you in some language other than English?				
Your father				
Your mother				
Your brothers or sisters				
Your grandparents				
Your playmates				
2. How often do you speak to any of these people in some language other than English?				
Your father				
Your mother				
Your brothers or sisters				
Your grandparents				
Your playmates				
3. How often does your mother speak to these people in a language other than English?				
Your father				
Your brothers or sisters				
Your relatives				
4. How often does your father speak to these people in a language other than English?				
Your mother				
Your brothers or sisters				
Your relatives				

PART II

1. Name the languages that are used in your home. __________ __________ __________

2. Can you read a language other than English? [ ] Yes [ ] No

3. About how old were you when you learned to speak English?

[ ] I Never did [ ] 2 to 4 years [ ] 5 to 7 years [ ] 8 or more years

4. About how old were you when you learned to speak a language other than English?

[ ] I Never did [ ] 2 to 4 years [ ] 5 to 7 years [ ] 8 or more years